NEURAL MECHANISMS OF MEMORY STABILITY AND FLEXIBILITY IN THE HUMAN BRAIN



Renée Suzanne Koolschijn
Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary Term 2022

ACKNOWLEDGEMENTS

Oxford has given me more than I could ever have imagined, and it is the people that have made my time here so worthwhile. Helen and Jill, I could not have asked for a better team of supervisors. Both of you still inspire me every day as role models of how to be great at science in your own way. Helen, it was a pleasure to spend so many hours together collecting data, tackling said data, and talking neuroscience in general. Thank you for training me up into the researcher I am today. Jill, thank you for making sure time and again I did not get lost in the details by keeping the big picture in mind. And thank you for even extending your help into the next phase of my career!

I am proud to have been part of the start of the amazing team Ginger. Even through most of our lab meetings were spent on zoom (thank you global pandemic) I really enjoyed our time together, and I am super excited to see what the lab will be up to in the near future.

Many people at WIN have made sure I will remember my time here very fondly. A few notable mentions: Alexa, thanks for providing me with a home when I was literally homeless, thank you to Zeena for helping me keep up with the WIN gossip, and Nicola for being an ONBI-buddy throughout this DTC-DPhil journey. Also a big thanks is due to everyone who helped to make the actual science happen, including the radiography team at WIN for running the MRI sessions (and Nicky in particular for lockdown tv recommendations), Hamed for everything RSA related, Uzay and Will for their spectroscopy support, Liliana and the BRC, and Mike and his team of medics for making noradrenaline a reality, Sue for proving much-needed friendly chat in a pandemic-struck FMRIB, and MANY people at WIN for helping out with all sorts of piloting.

A special shoutout goes to my childhood best friend Ellen: if it was not for you deciding to do a DPhil here, I would not never have thought about coming to Oxford. I figured I might as well see what Oxford was all about, and after a brief internship I was determined to pursue a DPhil here myself. Ellen, thank you for leading the way, and for being a friend throughout this journey!

Thank you to the rowers and coxes of OUWBC for helping me fulfil my athletic ambitions. Whilst spending so much time in a boat might not have been the most conductive to my DPhil research, from a life-experience point of view my time with you is irreplaceable. Thanks to The Oriel for adopting me as one of your own, it has been an absolute blast being part of this weird rowing cult. Alex, Dana, John, Becka and Jamie, I am so glad we crossed paths. Thanks to my friends Tom and Signe, and Kaitie and Patrick, for making me and Stephen not lose our sanity through the many lockdowns.

Hans, Inge, en HR, bedankt voor jullie steun en toeverlaat, en in het bijzonder voor het steunen van mijn beslissing om naar het buitenland af te reizen. Het is ontzettend fijn om al deze tijd in Engeland het gevoel te hebben gehad dat ik altijd terecht kon voor goede (haze)raad, of een luisterend oor als ik waar dan ook mee zat. Jackie, Gordon, and G Fran, thank you for making me feel part of your family, it was great to have a home to return to when Covid made travel abroad impossible.

Finally, and arguably more valuable than a Dr title, Oxford gave me a soulmate. Stephen, thank you for your support and love throughout, for dealing with all the stress that comes with trying to combine a DPhil with an almost full-time rowing schedule. It has been an unforgettable journey together so far, and I look forward to seeing what the future will bring us.

This DPhil has been supported by the UKRI Engineering and Physical Sciences Research Council (EPSRC) and Medical Research Council (MRC)-funded studentship (EP/L016052/1). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203139/Z/16/Z). The research presented in Chapter 4 was supported by the NIHR Oxford Health Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

ABSTRACT

The brain has the remarkable capacity to acquire, store, and reinstate memories. On the one hand, memories can be stored safely, free from interference. On the other hand, new memories are acquired daily, and can be used to update previously stored information. To safely store memories, a stable memory system is required, while for new learning flexibility and plasticity are needed. This means that the brain operates along a trade-off between memory stability and flexibility. In this thesis, I investigated the neural mechanisms that support stable and flexible memory, probing neocortical, hippocampal and neuromodulatory contributions. To study these different contributions, I have used techniques that measure and modulate cortical excitation and inhibition, in combination with behavioural and neural readouts of memory, with the aim to understand how neurochemistry and neural representations influence memory formation, storage and recall.

First, I asked whether neuromodulators might set the trade-off between stability and flexibility during learning by modulating the balance between excitation and inhibition (E/I). I manipulated the concentration of the neuromodulator noradrenaline during learning through administration of a single dose of the drug atomoxetine, and measured the effect of this intervention on cortical inhibition and behavioural and neural measures of overgeneralisation. I found that increased noradrenaline led to a decrease in neocortical inhibition, which predicted an increase in neural markers of excitability, which in turn predicted behavioural signatures of memory overgeneralisation.

Second, I went on to look at the mechanisms that protect stored memories from interference. By combining functional MRI (fMRI) with transcranial direct current stimulation (tDCS) to lower cortical inhibition, I showed that two distinct mechanisms prevent overlapping memories from interference: hippocampus separates overlapping memories using a relational code, while in neocortex, inhibition prevents co-activation of overlapping memories.

Third, I investigated the hippocampal and neocortical contributions to memory recall. By applying a novel MR sequence that interleaves fMRI with functional MR Spectroscopy (fMRS), I found that recall of a visual cue resulted in a transient increase in E/I balance in visual cortex, which was predicted by activity in hippocampus. Taken together, these three studies reveal the contributions of hippocampus and cortical E/I balance in memory-related processes, where hippocampus represents the relational structure of a memory and where neural inhibition gates

reactivation of memories and sets the trade-off between stability and flexibility during memory formation. Together, the hippocampus and cortical E/I balance play a key role in the neural computations that underlie flexible behaviour.

CONTENTS

List of figures ..... **ix**
List of tables ..... **xi**
Abbreviations ..... **xii**
1 Introduction **1**
1.1 Excitatory-Inhibitory Balance and Memory 2
1.1.1 Excitatory Plasticity 3
1.1.2 Inhibitory plasticity 3
1.1.3. Inhibitory interneurons 4
1.1.4 The role of inhibition in memory 8
1.1.5 Neuromodulation 11
1.2 Hippocampal mechanisms involved in memory 14
1.2.1 Anatomy of the hippocampal formation 15
1.2.2 Hippocampal indexing theory 16
1.2.3 Pattern separation 18
1.2.4 Pattern completion 19
1.2.5 Cognitive map theory 20
1.3 Hippocampal-neocortical interactions 21
1.3.1 Complementary learning systems 21
1.3.2 Memory-related interactions of hippocampus and prefrontal cortex 22
1.4 Thesis Overview 23
2 Methods for investigating brain acitivity and excitatory-inhibitory balance in humans **25**
2.1 Functional MRI 26
2.1.1 MRI fundamentals 26
2.1.2 The BOLD signal 27
2.1.3 fMRI preprocessing 29
2.1.4 fMRI univariate statistical analysis 32
2.1.5 fMRI multivariate statistical analysis 33
2.1.4 Repetition suppression 35
2.2 MR Spectroscopy 39
2.2.1 MRS physics principles 39
2.2.2 The MRS acquisition 41
2.2.3 MRS Analysis 44

2.3 Perturbing E/I balance 45
2.3.1 Transcranial Direct Current Stimulation 45
2.3.2 Pharmacological approaches 46
3 Time-resolved functional MRS: a tool to gain mechanistic insight into human cognition? 49
3.1 Introduction 50
3.2 fMRS task design: Blocked versus time-resolved 51
3.3 Measuring GABA: J-difference versus non-edited sequences 52
3.4 Analysis considerations concerning GABA quantification 53
3.5 Time-resolved fMRS in the literature 55
3.6 Interpreting fMRS-derived changes in metabolite concentrations 57
3.7 Concluding remarks 59
4 Distorting cognitive maps by elevating noradrenaline during learning 60
4.1 Abstract 61
4.2 Introduction 61
4.3 Methods 63
4.3.1 Participants and double blinding 63
4.3.2 Behavioural training 64
4.3.4 fMRI scan task 66
4.3.5 Pupillometry acquisition, preprocessing, and analysis 66
4.3.8 Univariate fMRI analysis and statistics 68
4.3.9 Assessing the relationship between fMRI, behaviour, and GABA+ 69
4.4 Results 69
4.4.1 Experimental design 69
4.4.2 Elevated noradrenaline during learning increases behavioural measures of memory distortions 72
4.4.3 Elevated noradrenaline modulates pupil response to surprising stimuli . 74
4.4.4 Elevated noradrenaline decreases inhibitory tone in neocortex 75
4.4.5 Hippocampus represents learned stimulus associations 76
4.4.6 Relating neural memory expression to behavioural distortions in the cognitive map 78
4.5 Discussion 80
4.6 Conclusion 84
4.7 Supplementary figures 85
5 The hippocampus and neocortical inhibitory engrams protect against memory interference 89

5.1 Abstract 90
5.2 Introduction 90
5.3 Methods 92
5.3.1 Participants 92
5.3.2 Behavioural training 92
5.3.3 fMRI scan task 95
5.3.4 fMRI imaging protocol 95
5.3.5 Magnetic Resonance Spectroscopy 96
5.3.6 Transcranial Direct Current Stimulation 97
5.3.7 Double-blind procedure for anodal/sham tDCS 98
5.3.8 fMRI data analysis 99
5.3.9 fMRI statistics and ROI specification 100
5.3.10 Representational Similarity Analysis 101
5.3.11 Correlations between fMRI, behavioural and MRS data 102
5.4 Results 102
5.4.1 Associative learning and experimental design 102
5.4.2 Hippocampus mediates memory interference using context-dependent relational codes 105
5.4.3 Manipulating neocortical E/I balance to measure the effect of inhibition on memory 109
5.4.4 Measuring associative memories using cross-stimulus suppression 111
5.4.5 Memory interference increases during periods of E/I imbalance 112
5.4.6 The interplay between the hippocampus and anterior LOC 116
5.4.7 Inducing behavioural memory interference using brain stimulation 116
5.5 Discussion 118
5.6 Conclusion 122
5.7 Supplementary figures 123
5.8 Supplementary Tables 131
6 Memory recall involves a transient break in excitatory-inhibitory balance 132
6.1 Abstract 133
5.2 Introduction 133
6.3 Methods 135
6.3.1 Participants 135
6.3.2 Virtual reality environment 135
6.3.3 Three-stage inference task 136
6.3.4 fMRI-fMRS scan task 138
6.3.5 fMRI-fMRS data acquisition 139

6.3.6 Trial categorisation during the inference test 140
6.3.7 fMRS metabolite quantification and analysis 141
6.3.8 MRS Monte Carlo simulations 143
6.3.9 fMRI preprocessing and GLMs 144
6.3.10 Univariate fMRI analysis and statistics 145
6.3.11 Assessing the relationship between fMRI and fMRS 145
6.4 Results 147
6.4.1 Task design and behaviour 147
6.4.2 Neural signatures of associative memory recall during inference 150
6.4.3 Dynamic increase in the ratio between glutamate and GABA in visual cortex during recall 152
6.4.4 A hippocampal index for fluctuations in neocortical ratio between glutamate and GABA 154
6.5 Discussion 155
6.6 Conclusion 161
6.7 Supplementary figures 162
6.8 Supplementary Tables 177
7 General discussion and future directions 181
7.1 Revealing the role of E/I balance in memory 181
7.2 Interactions between brain regions that support memory 182
7.3 Concluding remarks 183
Bibliography 184

LIST OF FIGURES

1.1 Locus Coeruleus and its projections in the human brain. 13
1.2 Anatomy of the human hippocampal formation. 16
1.3 Schematic of hippocampal indexing. 18
2.1 The haemodynamic response function. 29
2.2 Illustration of representational similarity analysis. 35
2.3 Schematic illustrating stimulus presentation design to measure XSS effects using fMRI BOLD. 38
2.4 Example 7T in vivo MRS spectrum 41
2.5 Illustration of GABA-edited MRS acquisition. 44
3.1 Monte Carlo simulations to assess the effect of including/removing prior constraints on GABA. 55
4.1 Experimental design 71
4.2 Noradrenergic stimulation during learning increases overgeneralisation 73
4.3 Elevated noradrenaline increases pupil surprise response 74
4.4 Elevated noradrenaline locally decreases inhibitory tone in neocortex 76
4.5 Hippocampus represents stimulus relationships and hippocampal activity predicts memory accuracy 77
4.6 Memory expression in hippocampus following learning under atomoxetine predicts later behavioural distortions in the cognitive map 79
4.7 Atomoxetine induces hippocampal coactivation between stimuli and unassociated but relationally proximal contextual cues 80
S4.1 Double-blinding results 85
S4.2 Inference test day 5 performance 85
S4.4 Noradrenaline does not induce changes in Glx and GABA+ in V1 87
S4.5 Coactivation between paired stimuli per group and group differences 87
S4.6 'Spread of association' XSS per group 88
5.1 Experimental design 104
5.2 The hippocampus mediates memory interference 107
5.3 Hippocampal representations show evidence for contextual pattern separation, organised by relational overlap 108

5.4 Manipulating neocortical E/I balance using brain stimulation unmasks otherwise dormant neocortical memories 111
5.5 Memory interference for more remote memories increases with brain stimulation 115
5.6 Relating learning accuracy to neural measures of memory interference 118
S5.1 Training task and surprise memory test 123
S5.2 Example of fMRI partial volume and RSA models with additional RSA analyses 125
S5.3 MRS spectra, subject inclusion, and additional analyses 126
S5.4 Replication of previous findings (Barron et al., 2016a) 127
S5.5 Simulated effect of tDCS across the brain and effect of cathodal stimulation on left prefrontal cortex (PFC) 129
S5.6 Memory interference effects in memory 1 and 2 130
6.1 Inference task design and behavioural performance 148
6.2 Behavioural performance in the inference test is predicted by performance on an associative test 149
6.3 Using fMRI-fMRS data to assess changes in BOLD signal and glu/GABA ratio during the inference test 151
6.4 Memory recall and inference involves a transient break in glu/GABA ratio 153
6.5 Hippocampal BOLD predicts neocortical glu/GABA ratio during recall 155
S6.1 Behavioural training and performance 162
S6.2 Comparison of different smoothing parameters applied to fMRI data 163
S6.3 Estimation of fMRS moving average 164
S6.4 Monte Carlo simulations to assess MRS data quality 165
S6.5 MRS data quality metrics across all spectra. 167
S6.6 The transient break in glu/GABA ratio observed during recall cannot be explained by changes in data quality metrics or goodness of model fit 168
S6.7 An increase in glu/GABA ratio in V1 during memory recall are also observed when categorising trials into 'remembered' and 'forgotten' using a less conservative approach 170
S6.8 The change in glu/GABA ratio is transient and only observed during memory recall 172
S6.9 The changes in metabolite concentrations cannot be attributed to changes in NAA:tCr 173
S6.10 During conditioning trials, no difference in glu/GABA ratio was observed 174
S6.11 Before and after memory recall, the hippocampal BOLD signal did not positively predict glu/GABA ratio in V1 175

LIST OF TABLES

S5.1 Average concentration of all metabolites measured using MRS 131
S6.1 The effect of sex on behaviour and on glu/GABA ratio in V1 177
S6.2 Number of trials per condition 177
S6.3 Number of trials split according to outcome 177
S6.4 fMRI contrast for 'remembered' – 'forgotten' 178
S6.5 Inter-subject covariance of glutamate and GABA 178
S6.6 Average number of spectra (NEX) 178
S6.8 Minimum Reporting Standards in MRS (MRSinMRS) checklist 179

ABBREVIATIONS

ASD	Autism spectrum disorder
ACC	Anterior cingulate cortex
ADHD	Attention-deficit hyperactivity disorder
ANOVA	Analysis of variance
ATX	Atomoxetine
BOLD	Blood oxygen level dependent
CA	Cornu ammonis
CHESS	CHEmical Shift Selective
Cr	Creatine
CRLB	Cramer-Rao Lower Bound
CSF	Cerebrospinal fluid
DABEST	Data Analysis with Bootstrap-coupled ESTimation
DG	Dentate Gyrus
E/I	Excitatory inhibitory
EC	Entorhinal Cortex
EEG	Electro-encephalography
EPI	Echo planar imaging
EV	Explanatory variable
FID	Free induction decay
fMRI	Functional magnetic resonance imaging
fMRS	functional magnetic resonance spectroscopy
FWE	Family-wise error
FWHM	Full-width at half-maximum
GABA	Gamma-aminobutyric acid
GLM	General linear model
Glx	Glutamate + glutamine
HLSVD	Hankel-Lanczos Singular Value Decomposition
HRF	Haemodynamic response function
IT	Inferior temporal
ITI	Inter-trial interval
LC	Locus coeruleus
LFP	Local field potential
LOC	Lateral occipital complex
LTD	Long-term depression
LTP	Long-term potentiation
MEG	Magnetoencephalography
MEGA	MEscher-GArwood

MM	Macromolecules
MNI	Montreal neurological institute
mPFC	Medial prefrontal cortex
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
MVPA	Multi-voxel pattern analysis
NAA	N-acetylasperate
NEX	Number of spectra
NMDA	N-methyl-D-aspartate
OC	Occipital cortex
PFC	Prefrontal cortex
PLC	Placebo
ppTMS	Paired-pulse transcranial magnetic stimulation
PTSD	Post-traumatic stress disorder
PV	Parvalbumin
RDM	Representational dissimilarity matrix
RF	Radiofrequency
ROI	Region of interest
RSA	Representational similarity analysis
SEM	Standard error of the mean
SNR	Signal-to-noise ratio
SOM	Somatostatin
SPM	Statistical parametric mapping
STDP	Spike-timing dependent potentiation
SVC	Small volume correction
tACS	Transcranial alternating current stimulation
tCr	Total creatine
tDCS	Transcranial direct current stimulation
TE	Echo time
TMS	Transcranial magnetic stimulation
TR	Repetition time
TUS	Transcranial ultrasonic stimulation
V1	primary visual cortex
VAPOR	VAriable Power and Optimised Relaxation delays
VIP	Vasoactive intestinal positive
VOI	Volume of interest
VR	Virtual reality
WM	White matter
XSS	Cross-stimulus suppression
ASD	Autism spectrum disorder

CHAPTER 1

INTRODUCTION

Throughout life, memories guide our everyday behaviour and decisions. They form the basis of who we are: our identity is shaped by our experiences and how we internalise them. But what are memories? The generally accepted view is that memories are like cortical reinstatements of experience: a memory is represented as an assembly of interconnected cells that activate as a unit as a result of strengthened excitatory synapses between them, as captured in Hebb's famous "neurons that fire together, wire together" postulate (Hebb, 1949). It follows then, that to form and store a memory, modifications to excitatory synaptic connections between neurons must be made to represent a new experience. Memory recall then represents the reactivation of these neural networks, which evoke the original experience.

While this account gives a basic understanding of memory, memory is more than a static repository of all experiences during our lifetime: memories are plastic and flexible, they are dependent on usage, and we are able to generalise and infer across past experiences. How does the brain implement these features?

Central to understanding the mechanisms through which brain manages memories is the trade-off between flexibility and stability. Memories need to be stored in a stable manner, safe from interference, to allow accurate recall. On the other hand, stored memories need to be editable so they can be adapted in accordance with new experiences, and mechanisms need to be in place that allow elements from different memories to be combined to support brain functions like inference and generalisation.

Brain architecture supports several mechanisms that enable these functionalities: In neocortex, the balance between excitation and inhibition (E/I balance) is thought to set the trade-off between stability and flexibility. For memory formation, inference, or generalisation to occur, E/I balance must be transiently broken to allow for synaptic modifications to be made. However, during this period of time memories may also be susceptible to interference. In the hippocampus, anatomy supports mechanisms like pattern separation, i.e. the orthogonalisation of experiences, as well as pattern completion. The trade-off between stability and flexibility is thus made explicitly visible in hippocampus: pattern separation promotes memory stability and protects from interference, while pattern completion promotes generalisation and inference.

In this thesis, I have investigated the neural mechanisms that support stable and flexible memory, probing neocortical, hippocampal and neuromodulatory contributions. In this introductory chapter I will cover the literature underlying the research questions I address in Chapters 4-6. The first section of this chapter covers the role of E/I balance in memory, specifically the role neural inhibition is thought to play, as this is a unifying theme throughout my DPhil research: in Chapter 4, I investigated whether the neuromodulator noradrenaline may mediate plasticity through modulating E/I balance. In Chapter 5, I hypothesised that for memories stored in neocortex, inhibition prevents overlapping memories from coactivation. In Chapter 6, I addressed the hypothesis that a perturbation of neocortical E/I balance may underlie memory recall. While my work is in humans and thus cannot directly uncover neural circuit-level mechanisms, my aim is to link my findings to work done in animal models where processes can be studied on a cellular level. Therefore I provide some background on inhibitory interneurons, the building blocks of these microcircuits. In the next section of this chapter, I summarise memory-associated processes and theories localised to the hippocampus, as all of my experimental work addresses interactions between hippocampus and neocortical E/I balance. This leads into the subsequent section, in which I cover the background literature on hippocampal-neocortical interactions, including complementary learning systems theory. Finally, in the last section I will give an overview of the experimental work presented in this thesis.

1.1 Excitatory-Inhibitory Balance and Memory

In the brain, new memories are thought to be incorporated into existing neural networks through changes in excitatory synapses. However, inhibitory synapses appear to play just as critical a role in memory, by balancing out the excitatory changes with inhibitory connections. These inhibitory changes contribute to the maintenance of E/I balance, and prevent run-away excitation. Furthermore, modulation of inhibition provides an elegant mechanism through which memory expression and the plasticity of neural networks could be regulated.

In this section, I will first introduce excitatory and inhibitory plasticity, and then discuss the role inhibition is thought to play in the different stages of a memory: i.e. formation, storage and recall. I will also give an overview of the more common inhibitory neuron classes and their proposed involvement in memory processing. A potential mechanism through which the brain could regulate (inhibitory) plasticity is through neuromodulation, and I will specifically discuss the neuromodulator noradrenaline, whose role in memory formation I investigated in Chapter 4.

Finally, I will briefly discuss the relevance of E/I balance in neuropsychiatric disorders, as anomalies in inhibition are thought to underlie various cognitive impairments related to memory.

1.1.1 Excitatory Plasticity

Memories are thought to arise from synaptic modifications made within networks of neurons. Synaptic plasticity is thus considered the mechanism that underlies new learning, as well as being the underlying pillar of the flexibility of our nervous system. But how do synaptic modifications arise? Most of our understanding of synaptic plasticity originates from the discovery of Long Term Potentiation (LTP) by (Bliss and Lomo, 1973) in the marine snail dentate gyrus. LTP refers to the activity-induced strengthening of synaptic efficacy, i.e. transmission efficacy between neurons increases when the pre- and post-synaptic neuron are active within a very short time interval (Bliss and Collingridge, 1993). In an experimental setting, LTP is typically induced by applying trains of high-frequency electrical pulses to the presynaptic axon. Similarly, Long Term Depression (LTD) arises when low-frequency stimulation is delivered to a neuron resulting in a decrease in synaptic transmission efficacy (Bliss and Lomo, 1973; Bliss and Collingridge, 1993; Kirkwood et al., 1993). LTP/LTD induction is thought to rely on N-methyl-D-aspartate (NMDA) receptor activation in the postsynaptic neuron, leading to an increase in Ca^{2+} (Nishiyama et al., 2000; Yang et al., 1999). Indeed, the properties of the postsynaptic NMDA receptor may underlie most forms of excitatory synaptic plasticity (Malenka and Nicoll, 1999; Urakubo et al., 2008).

LTP and LTD induction have been used as interventions to probe the nervous system in various species and brain regions and this methodology has led to a great amount of understanding of how excitatory plasticity can be induced and how these excitatory changes manifest in behaviour (Kirkwood et al., 1993; Malenka and Nicoll, 1999). However, it is still unclear how LTP and LTD relate to cognitive processes observed in vivo, like learning and memory (Cahill et al., 2001; Thompson, 2005).

1.1.2 Inhibitory plasticity

Excitatory and inhibitory currents within networks normally appear balanced, both locally and globally (Froemke et al., 2007; Okun and Lampl, 2008; Vogels and Abbott, 2009; Wehr and Zador, 2003). An example of matched inhibition occurs during sensory processing in neocortex, where presentation of a sensory stimulus invariably results in excitation followed immediately by inhibition (Gabernet et al., 2005; Wehr and Zador, 2003; Xue et al., 2014). Here, inhibition ensures temporally precise firing of neurons: there is a very brief window of opportunity of firing due to a

short delay between excitation and inhibition. Inhibition allows for the dynamic regulation of neural computations owing to the many different types of inhibitory interneurons that selectively synapse onto different regions of a targeted neuron (e.g. the axon, soma, dendritic regions) (Fishell and Rudy, 2011; Klausberger and Somogyi, 2008).

Given the requirement of E/I balance it follows that any modifications made to excitatory synapses must be matched by changes in inhibitory synapses in order to maintain E/I balance. Experimental observations and theoretical models confirm that, like their excitatory counterparts, inhibitory synapses also exhibit plasticity. In experiments on the rat auditory cortex, inhibitory tuning curves shifted over the course of several hours to follow induced changes in excitatory tuning curves (Froemke et al., 2007). Theoretical modeling has been applied to these experimental findings to demonstrate that inhibitory spike-timing dependent potentiation (STDP) can account for the changes in inhibitory tuning curves (Vogels et al., 2011). Indeed, it was later found in an in vitro study that STDP was required to rebalance neural circuits (D'amour and Froemke, 2015). However, the biophysical implementation of inhibitory plasticity is still poorly understood as the implementation and functionality of inhibition appears to vary per region and inhibitory cell type, of which there are many.

1.1.3. Inhibitory interneurons

In neocortex, inhibitory neurons make up about 20% of all neurons (Markram et al., 2004). Compared to excitatory neurons, inhibitory neurons are highly diverse and specialised in their connectivity and functionality. This supports the notion that inhibitory interneurons play an important role in the dynamic regulation of brain functioning, cognition and behaviour. Indeed, the computational functions of inhibitory neurons have been found to range from input separation through winner-take-all wiring schemes, context-dependent widening of neural response range, neuronal synchronisation, to controlling spike timing (Carvalho and Buonomano, 2009; Moore et al., 2010). Despite their diversity, inhibitory neurons do share a few common structural features which impact their functioning: they use GABA as their transmitter, most have aspiny/smooth dendrites and can receive both excitatory and inhibitory synapses on their soma (White, 1989). Their axons typically arborise laterally within a cortical column, and do not contact distant brain regions via white matter (Peters and Jones, 1984). The consequence of this is that interneurons in cortex mostly act locally rather than across a larger area. Inhibitory neurons are often referred to as inhibitory interneurons or simply interneurons, as they tend to connect other neurons with each other.

Inhibitory neurons can be classified using neuropeptide expression and calcium binding proteins, with the major classes being parvalbumin (PVs), somatostatin (SOMs) and vasoactive intestinal (VIPs) expressing inhibitory interneurons (Taniguchi et al., 2011). While these classes are by no means homogeneous, they are useful for gaining insight into neural computations as the classes are non-overlapping and cover ~85% of inhibitory cells (Rudy et al., 2011).

Parvalbumin positive interneurons

Characteristic properties of PVs are that they are fast-spiking and have low input-resistance (Kawaguchi et al., 1987), which gives them the ability to fire particularly rapid trains of action potentials to other neurons. PVs are highly interconnected which promotes synchronous activity (Galarreta and Hestrin, 1999). Two subtypes of PVs can be identified based on their morphology: basket cells that target the soma, and chandelier cells that target the axon initial segment (Kawaguchi and Kubota, 1997). Through targeting these sections of other neurons, PVs are particularly well positioned to have a controlling function by adjusting the gain of the integrated synaptic response in the targeted soma or modifying the action potential output in the targeted axon (Markram et al., 2004). They also are thought to be involved in more complex network operations, including pattern separation, modulation of place and grid cell field shapes, and gain modulation of sensory responses (Hu et al., 2014).

In hippocampus, PVs are associated with feedforward inhibition: by being optimised for fast firing, PVs can fire (and inhibit) before pyramidal cells or other interneurons do (Pouille et al., 2009). Here, feedforward inhibition narrows the integration window for excitatory potentials, enforcing coincidence detection and increasing temporal resolution. In both hippocampus and entorhinal cortex, PVs are also involved in feedback inhibition (Couey et al., 2013), which implements a winner-take-all computation (de Almeida et al., 2009). Here, the cell that manages to fire first will inhibit the other cells it synapses on. This might be an important computation contributing to orthogonalisation and pattern separation (Leutgeb et al., 2007), and may also underlie grid cell pattern firing (Pastoll et al., 2013).

PVs have been shown to regulate neural oscillations which play an important role in cognition and information processing across brain regions (Howard et al., 2003; Lundqvist et al., 2016). Suppression of PVs reduces gamma oscillations in cortex, while stimulation of PVs at gamma frequency leads to gamma oscillations in somatosensory cortex (Sohal et al., 2009). Furthermore, PVs have been shown to control theta oscillations in hippocampus (Losonczy et al., 2010).

Evidence specifying the involvement of PVs in plasticity and learning is increasing. For instance, following monocular deprivation, PVs in visual cortex are transiently inhibited which is thought to enable critical period ocular plasticity (Kuhlman et al., 2013). In auditory cortex, inhibition of PVs is necessary for auditory fear learning (Letzkus et al., 2011; Wolff et al., 2014). Finally, stimulation of PVs in prefrontal cortex accelerates extinction learning (Sparta et al., 2014). Thus, PVs are involved in various types of plasticity, with their computations depending on their connectivity within the local neuron network.

Somatostatin positive interneurons

SOMs comprise about 30% of all inhibitory cells in neocortex (Rudy et al., 2011) and are anatomically very diverse. Notable phenotypes include 'Martinotti type' low-threshold spiking (LTS) cells, regular spiking, and fast-spiking neurons (Ma et al., 2006). Cortical Martinotti cells target dendrites, especially those of pyramidal cells in layer I of cortex where they mediate inhibition between neighbouring pyramidal cells (Berger et al., 2009). In hippocampus, SOMs have been found to regulate burst firing, as opposed to PVs which were found to modulate phase (Royer et al., 2012).

SOMs have a low threshold for action potential generation, and in neocortex they exhibit high levels of spontaneous activity which is notably increased by the neuromodulators noradrenaline and acetylcholine (N. Chen et al., 2015; Gentet et al., 2012; Xiang et al., 1998). SOMs are densely wired into their local networks, both receiving strong inhibitory inputs as well as giving strong inhibitory outputs (Fino and Yuste, 2011). These properties, in combination with their tonic activity, make that they are in an excellent position to act as regulators of overall levels of inhibition throughout neocortex and hippocampus (Urban-Ciecko and Barth, 2016). For instance, up- and down regulation of SOMs spontaneous activity has been suggested to affect the effective connectivity between excitatory neurons and underlie changes in brain state (Urban-Ciecko and Barth, 2016). Like PVs, evidence supports the idea that SOMs integrate inputs from their local neuronal population. Compared to PVs, SOMs are more likely to be involved in providing modulatory inhibitory input as opposed to feedforward inhibition like PVs. This hypothesis is motivated by the differential inputs these different classes of interneurons receive (Murayama et al., 2009).

New learning has been suggested to be associated with up- and downregulation of basal firing rate of SOMs in a variety of paradigms, including fear learning and auditory discrimination tasks (Lovett-Barron et al., 2014; Pi et al., 2013). The observation that SOMs activity is reduced in V1

during associative visual learning (Makino and Komiyama, 2015) and during active sensation in general suggests the possibility that inhibition of SOMs facilitates synaptic plasticity and learning by gating excitatory transmission (Urban-Ciecko et al., 2015). After associative learning, SOMs increase their expression of GABA, suggesting that SOM-mediated inhibition may play a role in circuit plasticity following new learning, allowing the inhibitory network to adjust to learning-induced changes in the excitatory network.

In contrast, in a fear conditioning paradigm, presentation of the unconditioned stimulus was found to be associated with an increase in SOMs in hippocampus, which is thought to be in order to separate irrelevant sensory input from the aversive stimulus (Lovett-Barron et al., 2014). Therefore, while the involvement in learning of SOMs depends on the context and brain region, it is clear that SOMs exhibit significant plasticity and are more than simple relays for sensory information (Yavorska and Wehr, 2016).

Vasoactive intestinal positive interneurons

VIPs make up about 15% of all interneurons and are mostly located in superficial layers of cortex. Compared to other interneurons, they only contact with a small number of other neurons. Functionally, VIPs are thought to mediate disinhibition of pyramidal neurons, providing a form of gain control (Fu et al., 2015; Karnani et al., 2016a). VIPs preferentially target other interneurons, particularly SOMs, and to some extent also PVs (Pfeffer et al., 2013; Pi et al., 2013). As a result of their connectivity, VIPs are an excellent candidate for gating plasticity through disinhibition of SOMs. Indeed, activation of VIPs in primary visual cortex increases through inhibition of SOMs (Zhang et al., 2014), possibly opening transient windows for excitatory plasticity by releasing cells from the blanket of inhibition (Karnani et al., 2016b). This inhibitory microcircuit has been observed in auditory cortex and somatosensory cortex (Kuchibhotla et al., 2017; Lee et al., 2013), and has been associated with a range of cognitive processes, including associative learning and locomotion (Fu et al., 2014; Letzkus et al., 2011). Furthermore, in hippocampus CA3, amplified inhibitory drive of VIPs to PVs supported improved performance in a Morris water maze learning task in rodents (Donato et al., 2013).

However, as with the other inhibitory interneuron classes, the role of VIPs in neural microcircuits varies regionally and with the context of different behaviours. For instance, during locomotion-driven visual processing, both VIPs, SOMs and pyramidal cells have been reported to activate (Karnani et al., 2016a). In addition, other microcircuits enabling disinhibition not relying on VIPs have also been identified (Chamberland and Topolnik, 2012). Interestingly, VIPs are

capable of releasing not only GABA but also the neurotransmitter and modulator acetylcholine, indicating VIPs play roles other than disinhibiting local circuits (Dudai et al., 2020).

1.1.4 The role of inhibition in memory

Learning

To allow for plasticity accompanying new learning, it has been proposed that a break in inhibition is required. Indeed, theoretical and experimental studies have demonstrated that disinhibition is required for memory formation (Clopath et al., 2016; Letzkus et al., 2011). More precisely, a reduction of GABAergic tone is required to induce neocortical plasticity via LTP (Castro-Alamancos et al., 1995; Trepel and Racine, 2000). The role of disinhibition in animal models has been most extensively studied using auditory fear conditioning. Auditory fear conditioning paradigms typically involve the pairing of an initially neutral tone (the conditioned stimulus) to an aversive stimulus such as a foot shock (the unconditioned stimulus). Increasing GABAergic inhibition during learning interferes with spatial and aversive learning (Arolfo and Brioni, 1991; Brioni et al., 1989; Sanger and Joly, 1985). Similarly, mildly decreasing GABAergic inhibition results in improved learning (Brioni et al., 1989; Izquierdo et al., 1993). Additional evidence demonstrating the role of inhibition during learning comes from studies where disinhibition was found to be induced physiologically during memory formation. For instance, new learning is linked to a decrease in the expression of genes associated with inhibition such as GABA synthesising enzymes (Bergado-Acosta et al., 2008; Heldt and Ressler, 2007), and genes coding for several types of GABA-A receptors or associated molecules (Chhatwal et al., 2005; Heldt and Ressler, 2007).

In humans, most research on the role of inhibition in learning has been performed in studies investigating motor learning. Changes in inhibitory tone in humans are typically measured using Magnetic Resonance Spectroscopy (MRS, Chapter 2.2), or paired-pulse Transcranial Magnetic Stimulation (ppTMS). For instance, MRS-derived GABA was found to transiently decrease during motor learning in sensorimotor cortex (Floyer-Lea et al., 2006), and primary motor cortex (Kolasinski et al., 2019; Sampaio-Baptista et al., 2015) Similar outcomes have been observed when using ppTMS to probe intracortical inhibition during early phases of motor learning (Rosenkranz et al., 2007). Other studies have used an interventional approach where GABA is modulated using transcranial Direct Current Stimulation (tDCS) or a pharmacological intervention (Chapter 2.3). For instance, lowering cortical GABA in primary motor cortex by applying anodal tDCS improves motor learning (Nitsche et al., 2003; Reis et al., 2009), while increasing GABAergic inhibition using

the GABA agonist lorazepam disrupts motor skill acquisition (Bütefisch et al., 2000). Similar findings in relation to GABA in primary visual cortex suggest that this disinhibitory mechanism applies to regions other than motor and sensorimotor cortex as well: MRS-derived GABA was found to reduce in primary visual cortex in response to a period of monocular deprivation (Lunghi et al., 2015). However, it should be noted that a functionally opposite mechanism has been observed as well in both hippocampus and primary visual cortex. In the human visual cortex inhibition appears dependent on the task context, as MRS-derived GABA decreased during a target-detection task, but increased during a feature discrimination task (Frangou et al., 2019). In the mouse hippocampus dorsal CA1 region, in a contextual fear conditioning task, foot shocks have been found to activate SOM+ interneurons that suppressed activity in CA1 neurons (Lovett-Barron et al., 2014). A suggested explanation for this observation is that neural inhibition may be necessary to pattern separate the foot shock from the contextual fear memory: when hippocampal SOM+ interneurons are deactivated, the formation of the contextual fear memory is disrupted, indicating that this circuit is critical for proper encoding (Lovett-Barron et al., 2014). This idea is confirmed by neural network models that show inhibitory interneurons are required to control hippocampal activity during one-shot learning (Nicola and Clopath, 2019). Therefore, different dynamics of neural inhibition may reflect brain area and task-dependent processes.

Given the large variety in inhibitory neuron types, there has been an interest in identifying the contribution of different interneuron subtypes in learning. For this, animal models are the main option of investigation, and experiments using auditory fear conditioning have demonstrated that multiple types of interneurons are of importance during memory acquisition (Letzkus et al., 2015; Urban-Ciecko and Barth, 2016). In auditory and prefrontal cortex, the unconditioned stimulus activates VIP+ interneurons, which in turn project to PV+ and SOM+ inhibitory interneurons. This results in hippocampal pyramidal cells activating as they are released from inhibition (Letzkus et al., 2011; Pi et al., 2013). While it is not certain to what extent the mechanisms observed in these auditory fear conditioning paradigms apply to other types of learning, evidence suggests there are overlapping features between fear conditioning, episodic memory and multi-trial motor learning (Donato et al., 2013; Khan et al., 2018; Kuhlman et al., 2013). Therefore, this circuit mechanism for disinhibition may underlie the changes in GABAergic inhibition observed in humans during new learning.

Memory Storage

Following excitatory changes as a result of new learning, E/I balance needs to be maintained to ensure that neuronal networks do not become hyperexcitable as a result of the changes.

Theoretical and experimental evidence suggest that inhibitory plasticity plays an important role in restoring E/I balance by matching the changes in excitation with inhibition (D'amour and Froemke, 2015; Hennequin et al., 2017; Vogels et al., 2011; Xue et al., 2014). Besides inhibitory rebalancing, it is worth mentioning that other homeostatic processes like synaptic scaling are also thought to contribute to cortical stability (Turrigiano, 2008). Synaptic scaling refers to the normalisation of all excitatory synapses within a neural network.

Inhibition may act both locally and globally in order to stabilise neural networks after learning. Studies in rodents indicate that, like the excitatory changes as a result of learning, inhibitory changes are tuned to the relevant stimulus (D'amour and Froemke, 2015; Froemke et al., 2007). Theoretical models illustrate how tuned inhibition is necessary to prevent spontaneous memory activation (Linkerhand and Gros, 2013; Litwin-Kumar and Doiron, 2014; Vogels et al., 2011; Zenke et al., 2015). Electrophysiological recordings in songbirds are consistent with this model (Vallentin et al., 2016). While zebra finches are young, excitatory premotor neurons are active as they listen to a tutor's song. During adulthood, these premotor neurons are silent, but they can be released from inhibition via application of a GABA agonist, gabazine, which disinhibits the neural circuit and produces a similar neural response as observed in juvenile birds. Furthermore, during learning, as juvenile birds improve in their ability to imitate a tutor's song, inhibitory currents in the premotor area become more precisely tuned to the song (Vallentin et al., 2016).

Results of a human neuroimaging study also confirm the idea that neural inhibition matches excitatory changes accompanied by new learning: immediately after learning, expression of the newly acquired memories was initially detected using fMRI adaptation (Chapter 2.1) but this expression decreased over time (Barron et al., 2016a). At a later point in time, when cortical inhibition was lowered using anodal tDCS, re-expression of dormant cortical memories was detected (Barron et al., 2016a), mirroring behaviour observed in a neural network model (Vogels et al., 2011). These results are in line with the idea that inhibitory plasticity silences new excitatory networks shortly after learning takes place, storing memories in a dormant state.

Memory Retrieval

Recall of a memory is thought to occur through a reinstatement of the cortical activity, likely coordinated by a hippocampal index (Horner et al., 2015; Pacheco Estefan et al., 2019; Tanaka et al., 2014) (Chapter 1.2.2). Within the framework suggested in the previous paragraphs, where inhibition keeps memories in a dormant state, disinhibition would therefore be a plausible mechanism through which a memory could be reinstated. Theoretical models corroborate this

hypothesis by showing that retrieval can be induced by transiently modulating synaptic inputs, effectively perturbing E/I balance (Vogels and Abbott, 2009). In addition, auditory fear conditioning studies in rodents show that presentation of the conditioned stimulus induces strong, phasic inhibition in a subset of inhibitory interneurons in dorso-medial prefrontal cortex, but only when the rodent exhibits the conditioned fear response (Courtin et al., 2014). Furthermore, this disinhibition is both required and sufficient for memory re-expression, as optogenetic inhibition of the same inhibitory interneurons induced the fear response (Courtin et al., 2014). These findings suggest that disinhibition may underlie both memory acquisition as well as memory retrieval.

What mechanism could mediate disinhibition during memory reinstatement? While this remains an open area of research, at the cellular level neuromodulators could be involved in regulating this process, for instance through cholinergic projections from the basal forebrain (Kim et al., 2016). On a brain region level, hippocampus and prefrontal cortex could play a role in mediating disinhibition in cortical areas as both these regions are thought to be involved in separating context-dependent memories (Bulkin et al., 2016; Depue et al., 2007).

1.1.5 Neuromodulation

Only a fraction of the inputs the brain receives result in long-term synaptic changes and thus long-term memories (otherwise our brain would soon be overflowing with full accounts of every detail of every day of our life!). What neural mechanisms drive when and where in the brain neural plasticity occurs in vivo?

Neuromodulation has been identified as a major mechanism through which neural plasticity can be guided (Froemke and Martins, 2011; Gu and Yakel, 2011; Pawlak et al., 2010). Numerous neuromodulators have been identified that affect synaptic plasticity in neural circuits. Indeed, blocking the activity of neuromodulators like acetylcholine or noradrenaline can prevent associative and perceptual learning (Fletcher and Wilson, 2002; Kroon and Carobrez, 2009). Behaviourally, neuromodulators are particularly associated with attention. It has been suggested they signal salience in response to uncertainty or unexpected situations, requiring the updating of beliefs about the world (Yu and Dayan, 2005). This framework suggests that uncertainty triggers the release of neuromodulators which 'select' the experiences to be committed to memory or updated (Hasselmo, 2006).

Consistent with these ideas is the growing body of work indicating that neuromodulators can act by disinhibiting neural networks, thereby transiently disrupting E/I balance to allow excitatory changes to be made. For example, stimulation of the cholinergic system has been shown to lower

feedforward inhibition (Froemke et al., 2007; Metherate et al., 1992; Woody and Gruen, 1987), and evidence suggests that acetylcholine release facilitates LTP and LTD in hippocampus (Gu and Yakel, 2011; Sugisaki et al., 2011). The neuromodulator noradrenaline is associated with disinhibition (Kuo and Trussell, 2011), and its role in memory is evidenced by studies showing a reduction or blocking of noradrenaline and its receptors results in memory impairments (Murchison et al., 2004). As a last example, dopamine has long been associated with reward and prediction errors, and has been shown to regulate both excitation and inhibition, depending on the type of receptor that is targeted (Seamans and Yang, 2004).

It should be noted that these different neuromodulatory systems do not work independently but instead are all directly connected: Locus coeruleus, the main source of noradrenaline in the brain, receives dopaminergic input from the ventral tegmental area and substantia nigra, as well as serotonergic input from the raphe nuclei of the midbrain. From locus coeruleus, noradrenergic projections target ventral tegmental area, substantia nigra and the basal forebrain, the main source of acetylcholine in the brain. In addition to noradrenergic inputs, the basal forebrain also receives dopaminergic and serotonergic projections (Eggermann et al., 2001; Sara and Bouret, 2012).

Noradrenaline

Noradrenaline or norepinephrine is a neuromodulator strongly associated with attention and salience (Aston-Jones and Cohen, 2005; Eldar et al., 2013; Sara, 2009). Noradrenaline is released through noradrenergic projections originating from Locus Coeruleus (LC, Latin for "blue spot"), located in the posterior area of the rostral brainstem near the fourth ventricle (Figure 1.1). LC neurons respond to novel/surprising stimuli of any type (visual, auditory, sensory) by phasically discharging (Aston-Jones and Bloom, 1981; Foote et al., 1980).

To understand the function of noradrenergic neuromodulation, it is helpful to compare its effects to cholinergic neuromodulation: while both noradrenaline and acetylcholine are involved in attention, they likely modulate cortical processing and behaviour in different ways as a result of the differences in organisation of their projections. Cholinergic projections from nucleus basalis project more strongly to the cortex, while LC projects throughout the entire central nervous system (Figure 1.1) (Berridge, 2008; Fuxe et al., 2012).

It has been suggested that acetylcholine release is driven by 'expected uncertainty', while noradrenaline is released upon 'unexpected uncertainty' (Yu and Dayan, 2005). Unexpected uncertainty arises e.g. when a stimulus previously reliably lead to an outcome and suddenly becomes unreliable and leads to unexpected outcomes. These changes in the world require

updating of neuronal circuits via synaptic plasticity, and it is proposed that noradrenaline facilitates this through disinhibition. This hypothesis is supported by studies investigating the effect of LC stimulation on sensory-evoked responses. For instance, repetitively pairing a pure tone with LC stimulation dramatically increases responses in auditory cortex. In fact, after initial LC stimulation responses to all stimuli (whether paired with stimulation or not) is increased (Edeline et al., 2011; Martins and Froemke, 2015). In vitro and in vivo experiments in rodents suggests that this is due to a decrease in tonic inhibition in response to LC stimulation, as opposed to phasic, stimulus-evoked inhibition observed with cholinergic stimulation (Kuo and Trussell, 2011; Martins and Froemke, 2015). The result of this tonic disinhibition is greatly increased excitability in response to any input. In these experiments, after ~90 minutes excitability returns to normal levels and the neurons respond only to the stimulus that was combined with the LC stimulation. Thus, LC stimulation results in increased gain between neural input and response, and this broadband enhancement of responsivity is suggested to be a mechanism for noradrenaline to sensitise neural populations to inputs during times of unexpected uncertainty (Froemke, 2015).

The longevity of the synaptic changes induced by LC stimulation is not widely investigated yet, but changes in neural tuning in response to LC stimulation persist for at least several hours or longer (Martins and Froemke, 2015). In addition, after initial pairing, presentation of the sensory stimulus alone can reactivate LC, resulting in additional release of noradrenaline, which would be a mechanism by which the changes in neural circuits could be maintained. However, it should be noted that long-term changes are more selective when instead of LC stimulation noradrenaline iontophoresis is used to induce plasticity (Manunta and Edeline, 2004), where iontophoresis is arguably a manipulation closer to natural neuromodulator release.

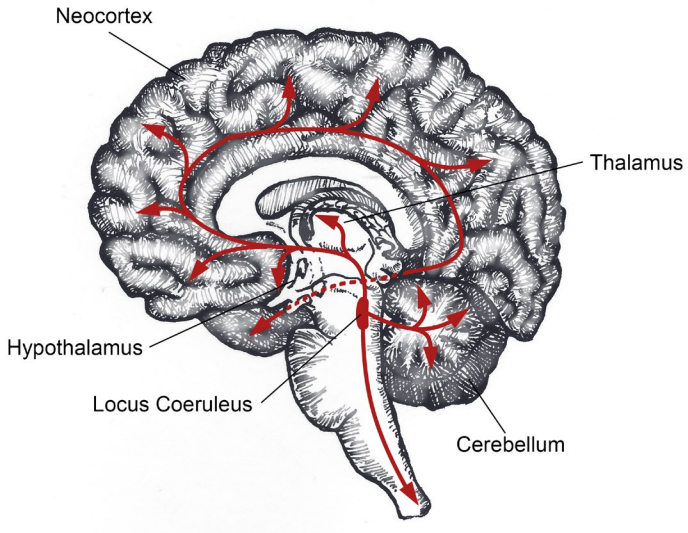


Figure 1.1 Locus Coeruleus and its projections in the human brain.

1.1.6 Clinical relevance of E/I balance

The importance of E/I balance for memory and cognition in general is underlined by observations that many cognitive impairments are associated with disturbances in E/I balance. E/I balance may act to suppress spontaneous activation, mask unwanted perception and inhibit unwanted behaviour. Maladaptive inhibition or perturbations in E/I balance have been suggested to underlie conditions ranging from schizophrenia (Lisman et al., 2008; Vogels and Abbott, 2007), autism spectre disorder (ASD) (Gogolla et al., 2009; Nelson and Valakh, 2015) post-traumatic stress disorder (PTSD) (Patel et al., 2012) to epilepsy (Peñagarikano et al., 2011). For instance, schizophrenia is associated with hippocampal hyperactivity, attributed to a deficit in GABAergic inhibition (Heckers and Konradi, 2015). In the case of PTSD, low levels of plasma GABA level immediately following the traumatic event were predictive of individuals' tendency to develop PTSD, with lower levels of GABA being associated with higher incidence of PTSD (Vaiva et al., 2004). Particularly for conditions associated with intrusive thoughts or behaviour, like schizophrenia, PTSD and ASD, disturbances in inhibitory mechanisms provide a compelling explanation of the symptoms observed. For instance, an associative memory not properly balanced by inhibition would be predicted to be unusually strong and vivid, which is a symptom associated with ASD (Zamoscik et al., 2016), and could underlie delusional experiences and hallucinations linked to schizophrenia (Vogels and Abbott, 2007).

Further in support for the hypothesis that E/I imbalance underlie cognitive impairments are studies on cognitive control, where subjects are tasked to control unwanted thoughts. Here, increased ability to control their thoughts was found to be associated with a higher hippocampal concentration of MRS-derived GABA (Schmitz et al., 2017).

1.2 Hippocampal mechanisms involved in memory

The hippocampus has been studied extensively in relation to its role in learning and memory. One of the key reasons for this is the case of patient H.M., who after removal of his bilateral hippocampi and its surrounding regions lost the ability to encode new memories (Scoville and Milner, 1957). Later experiments in animals confirmed the importance of hippocampus in memory using more specific genetic or pharmacological manipulations (Kim and Fanselow, 1992; Morris and Frey, 1997). Another reason for the interest in hippocampus in the context of memory is the pioneering work by O'Keefe and colleagues in the 1970s, who discovered cells in the hippocampus which fired according to an animal's physical location in space, aptly named 'place cells' (O'Keefe

and Dostrovsky, 1971). This work provided the first experimental evidence for the theory that the brain constructs cognitive maps of the world, as hypothesised by Tolman (Tolman, 1948), and identified the hippocampus as the location where these cognitive maps could be represented. Cognitive maps are thought to underlie flexible behaviour, enabling us to take mental shortcuts and giving us awareness of the underlying structure in the world.

The hippocampus is part of the hippocampal formation, located in the medial temporal lobe and plays a key role in the processing of experiences and memories. Within the framework of the hippocampal-cortical memory system, the hippocampal formation is often presented as providing an index for memories stored in the cortex.

In this section, I will cover the two main theories that aim to describe the function of hippocampus in memory, namely the hippocampal indexing and cognitive map theories. As the indexing account is rooted in the hippocampal anatomy, I will first review the structure of the hippocampal formation in humans before discussing the neural processing mechanisms involved in memory linked to this region of the brain.

1.2.1 Anatomy of the hippocampal formation

The hippocampal formation is commonly defined as the hippocampus itself and the surrounding substructures that together form a functional unit involved the processing of memory: i.e. the Entorhinal Cortex (EC), the Dentate Gyrus (DG) and the subiculum (Figure 1.2) (Andersen et al., 1971). The hippocampus itself can be further divided into 3 subregions, the Cornu Ammonis 1 (CA1), CA2 and CA3 regions. The hippocampal formation features a unique unidirectional connectivity which sets it apart from the rest of the brain, and it is thought this connectivity underlies its memory processing capabilities.

The EC acts as a terminal to the other hippocampal structures, receiving highly processed information for neocortex: it collects its inputs from projections across the neocortex, including perirhinal cortical areas, parahippocampus, prefrontal cortex (PFC) and association cortices (Canto et al., 2008). Different subregions of the EC then project to the granule cells in the DG, CA3 and CA1. Granule cells in the DG subsequently connect to CA3 pyramidal neurons via mossy fibers. A single CA3 neuron receives very strong inputs from relatively few mossy fibers (<50 inputs), which would be a mechanism to orthogonalise inputs: this connectivity is thought to underlie pattern separation (Section 1.2.3).

Pyramidal cells in CA3 have extensive recurrent excitatory connections to other neurons in CA3. Another feature of CA3 is high sensitivity to LTP, and these two properties result in CA3

acting as an autoassociative pattern completion network (Chapter 1.2.4). CA3 neurons project onwards to CA1, the region where place cells, considered integral to instantiations of cognitive maps, are located (Chapter 1.2.5). Finally, CA1 neurons project to the subiculum and to layer V of the EC, completing a unidirectional hippocampal loop. The subiculum is considered the main output area of the hippocampal formation as it connects widely to neocortex, as well as closing the hippocampal loop by projecting to EC (O'Mara et al., 2001). Outward projections of layer V of EC mirror the incoming afferents to EC, projecting to neocortical areas via parahippocampus and perirhinal cortex.

CA1, subiculum and the EC also have dense reciprocal connections with PFC, a region associated with the contextual aspect of memories and cognitive control (Bunsey and Eichenbaum, 1996; Koechlin et al., 2003). Medial PFC and Orbitofrontal cortex connect back to the hippocampal formation via the entorhinal cortex, parahippocampus and the perirhinal cortex. These reciprocal connections enable a dialogue between the two regions (Chapter 1.3.2).

Finally, the hippocampus also receives subcortical inputs from many other regions, including amygdala, thalamus, raphe nuclei and locus coeruleus. These inputs may account for neuromodulator modulated effects on memory.

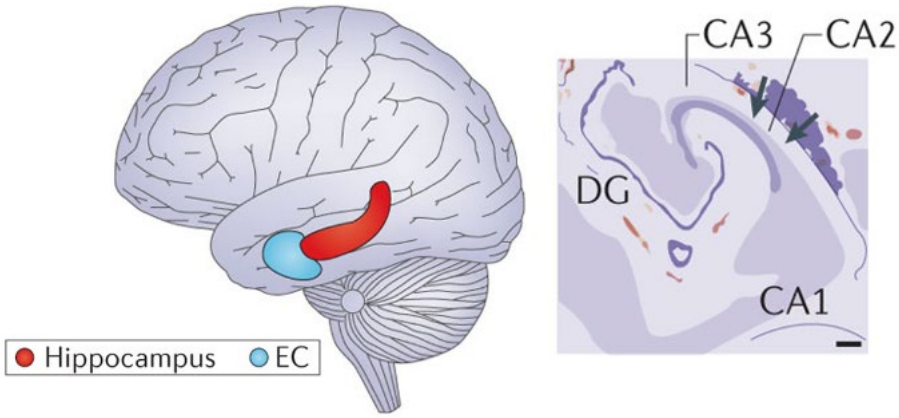


Figure 1.2 Anatomy of the human hippocampal formation. Adapted from (Strange et al., 2014).

1.2.2 Hippocampal indexing theory

The hippocampal indexing theory was proposed by Teyler and Discenna (1986), and states that the hippocampus forms and contains indices that give rapid and efficient access to memory engrams stored in neocortical areas (Teyler and DiScenna, 1985). The theory accounts for memory encoding, storage and retrieval of experience by separating memory into two conceptual parts: the stored representation of the memory and reproduction of the activity pattern. Hippocampal

indexing provides the framework for the complementary learning systems account (McClelland et al., 1995) (Chapter 1.3.1).

Figure 1.3 illustrates the proposed working mechanism of the hippocampal index. An experience is represented by the brain as a pattern of activity in neocortex. This pattern of activity projects to the hippocampus, and this conjunctive firing strengthens the connection between the neocortical neurons and the coactivated hippocampal neurons (Hebb, 1949). The resultant is a set of strengthened synapses in hippocampus, the memory trace: a hippocampal representation that links to an activity pattern distributed over neocortex. During retrieval, a portion of the neocortical pattern is reactivated. Signals are sent through a part of the memory trace, and these projections activate the hippocampal neurons representing the original and complete experience. The hippocampal neurons then project back to neocortex to activate the pattern representing the complete memory. This illustrates the idea that the hippocampus does not contain the memory itself, but it provides and index to elsewhere in the brain that allows reactivation of the episode.

The theory gives a possible account for ongoing plasticity, i.e. as connections between hippocampus and neocortex are the result of Hebbian plasticity, repeated activation will make the memory trace stronger even if the activation is offline, e.g. during sleep. This fits in well with the experimental findings around hippocampal replay (Chapter 1.3.1).

Hippocampal indexing theory is supported by the anatomy and physiology within the hippocampal formation and its connections to other brain regions. The hippocampus, through EC and parahippocampal cortex, receives inputs from many neocortical regions and would therefore be capable of combining different sensory features that make up a behavioural episode. Through the hippocampal loop, the hippocampal then projects back to neocortex and thus targets the same neocortical areas.

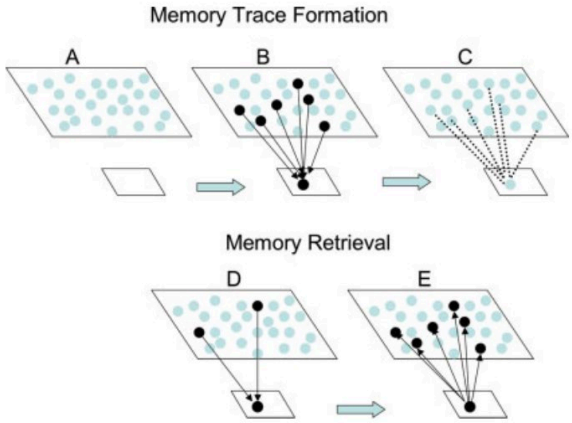


Figure 1.3 Schematic of hippocampal indexing. A-C Formation of a hippocampal index. A neocortical activity pattern in (B) projects to hippocampus and a lasting index is formed (C). **D-E** Memory retrieval. A subset of the original input reactivates the hippocampal index in (D), and upon activation the hippocampus projects back to neocortex, completing the original activation pattern. Adapted from (Teyler and Rudy, 2007).

For hippocampal indexing theory to be conceivable, some form of rapid plasticity would need to be available in hippocampus to rapidly generate the inwards and outwards memory traces based on a single experience (Insausti et al., 1997; Swanson and Kohler, 1986). Relatively modest amounts of afferent input to hippocampus has been shown to engage NMDA-receptor dependent Ca^{2+} influx, which rapidly induces LTD (Collingridge et al., 1983) that can easily be reversed (Wagner and Alger, 1996). This would be an excellent candidate mechanism for rapidly generating an index.

1.2.3 Pattern separation

A crucial aspect of memory is the ability to discriminate among similar experiences, which the hippocampus is thought to support through pattern separation. Pattern separation refers to the process of storing similar representations in a distinct, non-overlapping or orthogonalised manner (Yassa and Stark, 2011).

DG granule cells are thought to underlie pattern separation in hippocampus, owing to the fact that their sparse but strong mossy fibre projections to CA3 would be a way to implement orthogonalisation during learning (Ribak et al., 1985). Lesion studies indeed show that DG is required for spatial pattern separation and blocking DG input to CA3 impairs encoding (Gilbert et al., 2001; Lee and Kesner, 2004). Furthermore, even with small changes in input, DG was shown to separate stimulus representations (Leutgeb et al., 2007). Neurogenesis in DG has also been identified as playing a functional role in pattern separation (van Praag et al., 2002). Preventing dentate neurogenesis impaired learning in a contextual fear discrimination task (Tronel et al., 2012). However, the exact mechanism is still a matter of debate (Yassa and Stark, 2011).

There is also evidence in humans for pattern separation in hippocampus, measured using multivariate fMRI approaches: low representational similarity was observed between familiar stimuli and lures specifically in the DG/CA3 area, while other hippocampal regions showed overlapping representations for stimulus and lure (Bakker et al., 2008). In another study, activity patterns in DG were used to successfully classify similar scenes, supporting the role of DG in pattern separation (Berron et al., 2016). Finally, using pattern similarity analyses, Favila and colleagues showed that the extent of representational overlap at encoding predicted later memory interference, providing an interesting account where pattern separation could be linked memory interference (Favila et al., 2016).

1.2.4 Pattern completion

An integral part of the hippocampal indexing theory is the proposed mechanism for reactivation of a memory trace: a subset of the experience that originally established the index must lead to the retrieval of the complete representation (Figure 1D,E). This process is referred to as pattern completion. Theoretical models indicate that CA3 with its many recurrent connections enables it to act an auto-associative pattern completion network via attractor dynamics, making it particularly suitable for this task (Hopfield, 1982; Treves and Rolls, 1994). Pattern completion also supports generalisation as follows: multiple similar (incomplete) cues could trigger reactivation of the entire memory trace (Rolls, 2008). It is worth noting that pattern completion, while facilitating recall, may also be a source of memory interference or overgeneralisation if representations in CA3 are overlapping.

Evidence for CA3 pattern completion comes from experiments where recall is studied, where pattern completion would be of critical importance. Rodent studies have demonstrated that CA3-lesioned mice show impaired recall of object-location associations and impaired performance in a Morris water maze task when only partial cues were given (Gold and Kesner, 2005; Nakazawa et al., 2002).

In humans, evidence for pattern completion is assessed in a manner analogous to animal models: a partial cue related to a more complex memory is presented and reinstatement of the whole event is measured. Reinstatement can be measured using fMRI through elegant task designs in combination with computational modelling or advanced analysis techniques like multi-voxel pattern analysis (MVPA). Using an ROI-based approach, Chen et al. (2011) found Blood oxygen level dependent (BOLD) signal in CA2/CA3/DG predicted recall of a paired stimulus (Chen et al., 2011). Using an MVPA-based approach, Shapiro et al. (2012) found that CA2/CA3/DG

predicted upcoming stimuli that were part of a temporal sequence (Schapiro et al., 2012). Finally, using fMRI BOLD and computational modelling, a reinstatement effect was found in neocortex which was predicted by hippocampal BOLD, supporting the idea that activity in hippocampus (i.e. pattern completion) mediated cortical reinstatement (Horner et al., 2015). Furthermore, fMRI has also revealed evidence supporting the presence of attractor dynamics in hippocampus, supporting theoretical models proposing that CA3 indeed acts as an auto-associative completion network (Steemers et al., 2016). While the resolution of fMRI still limits finer spatial localisation of memory-supporting processes like pattern completion and separation, results in fMRI studies do support theoretical models that propose the hippocampus supports pattern completion.

1.2.5 Cognitive map theory

How are humans and other animals able to make inferences using partial information, and rapidly adapt their behaviour in response to new information? Tolman (1948) proposed that these capabilities could be achieved through the brain forming cognitive maps of the world, where relational information is organised in a map-like fashion (Tolman, 1948). Such a mental map would be a very powerful tool to allow flexible behaviour, for instance one could imagine traveling through an abstract memory space of previous experiences to make decisions in novel situations (Eichenbaum and Cohen, 2014).

Tolman was led to the idea of a cognitive map through his experiments in which he studied rats learning to navigate a maze. While Tolman's notion of the cognitive map went beyond physical space into all domains of behaviour, it has mostly been applied to spatial navigation. This may be due to the discovery that such a cognitive map of physical space is indeed represented in the hippocampus by CA1 neurons called 'place cells' (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadel, 1978).

Place cells track the activity of an animal during spatial navigation, and fire as the animal occupies a specific location within an environment (Wills et al., 2010). They also carry information on the location of behaviourally relevant objects and events, and therefore are relevant for memory (Komorowski et al., 2009). Place cells fire independently of the direction an animal is moving, i.e. they represent an allocentric map (Leutgeb et al., 2005; O'Keefe and Nadel, 1978). Another type of neuron found in the medial EC, called grid cell, is also thought to be critically involved in the neural representation of cognitive maps. These cells are the most common cell type in medical EC and characteristically fire at multiple place fields equally spaced on a triangular grid (Hafting et al., 2005; Sargolini et al., 2006). Through this property, they can represent relationships and distances

between different spatial locations (Bush et al., 2015; Stemmler et al., 2015). Whilst place and grid cells have been most extensively studied using animal models, there is also direct evidence for the presence of place and grid cells in the human hippocampus and EC respectively, for instance from recordings in epilepsy patients (Ekstrom et al., 2003; Jacobs et al., 2016).

Increasingly, evidence confirms the notion that place and grid cells do not only encode spatial dimensions, but also non-spatial maps. For instance, hippocampus has been found to map time, by 'time cells' that fire at sequential moments in temporally ordered experiences (MacDonald et al., 2011; Pastalkova et al., 2008; Paz et al., 2010). Results from these cellular recordings are corroborated by human neuroimaging studies showing that hippocampus encodes temporal information during learning and retrieval of stimulus sequences (Hsieh et al., 2014; Kalm et al., 2013). Furthermore, human neuroimaging work has also provided evidence that cognitive maps in human hippocampus/EC span the dimension of social space (Tavares et al., 2015), stimulus transition space (Garvert et al., 2017; Schapiro et al., 2013) or an arbitrary non-spatial dimension like bird-space (Constantinescu et al., 2016).

It is worth noting that hippocampal indexing and cognitive map theory are not mutually exclusive, and multiple theories exist on how the two can be exactly reconciled, e.g. (Buzsáki and Moser, 2013; Eichenbaum and Cohen, 2014; Whittington et al., 2020).

1.3 Hippocampal-neocortical interactions

A central aspect of hippocampal indexing theory is the separation of a memory index in the hippocampus and the memory traces in neocortex. This implies that there must be an interaction between hippocampus and cortical areas, and several theories have been proposed how this interaction might function. The most influential one to date has been complementary learning systems theory (McClelland et al., 1995). In this section, I will introduce this theory and briefly cover hippocampal replay, a process thought to underly hippocampal-neocortical consolidation. Finally, I will cover the relationship between hippocampus and PFC, which is considered important for processing the contextual element of memories.

1.3.1 Complementary learning systems

The complementary learning systems theory posits that effective learning requires two memory systems: one, located in hippocampus, rapidly learns the specifics of individual episodes, and another, located in neocortex, serves as a storage for structured knowledge about the environment and the world. This framework, developed by McClelland and colleagues (1995) and which

incorporated ideas and models of the function of hippocampus and neocortex (Marr, 1971; Sherry and Schacter, 1987; Larry R. Squire, 1992; Treves and Rolls, 1994), is based on the trade-off between two opposing computational goals: learning about specifics versus generalising. Theoretical models illustrate that to accomplish both, two memory systems would be required as one system would be prone to catastrophic interference where memories get overwritten and are forever lost (McCloskey and Cohen, 1989).

The idea that hippocampus is a specialised region for memorising patterns of activity in neocortex that can later be reactivated fits well with accounts of hippocampal indexing, pattern completion and separation, while overlapping and long-lasting representations of memories in neocortex would be a mechanism for generalisation based on previous experience. Later experimental observations have proved that some aspects of the theory were not accurate and so the theory has been updated over time (Nadel and Moscovitch, 1997; Wang and Morris, 2010), e.g. to incorporate observations that cortex can act both as a consolidating system as well as a fast-learning system (Tse et al., 2007). Still, the principle of 'division of labour' between hippocampus and cortex holds.

Consolidation of memories from hippocampus to cortex is thought to rely on the reactivation of activity patterns, or replay, of the neural activity that occurred during the experience (Louie and Wilson, 2001; Wilson and McNaughton, 1994). There is an abundant number of studies supporting replay of place cell firing sequences both during periods of sleep (Dupret et al., 2010; Ji and Wilson, 2007) as well as during periods of quiet wakefulness (Karlsson and Frank, 2009). Replay is thought to occur through sharp wave ripples (SWRs), where neural sequences during wakefulness are compressed into brief firing sequences of tens of milliseconds (Karlsson and Frank, 2009; Nádasdy et al., 1999). Interestingly, replay in rodents and coordinated SWRs between hippocampus and cortex in humans has also been linked to memory retrieval (Singer et al., 2013; Vaz et al., 2019).

1.3.2 Memory-related interactions of hippocampus and prefrontal cortex

The interaction between hippocampus and PFC deserves special attention: PFC is considered particularly involved in representing the contexts of memories, rather than representing sensory information like many other cortical areas. This interaction could therefore be critically important in preventing interference or unwanted generalisation between similar or related experiences. The proposed functional role of PFC is often presented as a railway track metaphor: during memory formation, hippocampus lays down new tracks, while PFC is subsequently responsible for switching between tracks depending on the situation or context (Miller and Cohen, 2001). During

memory retrieval, PFC would then bias reactivation in hippocampus based on the contextual cues presented. The interaction is supported through the direct projections from (human anterior) hippocampus and PFC, and connections from PFC back to perirhinal and EC (Navawongse and Eichenbaum, 2013). Notably, PFC can be subdivided into several functionally distinct areas, and it appears medial PFC (mPFC) is particularly involved in memory-related processes.

Experimental evidence of the role of PFC in guiding recall of contextual-contingent information comes from rodent lesion studies, which show that damage to PFC impairs the ability of animals to switch between different types of memories (Ragozzino et al., 2003; Rich and Shapiro, 2007). Specifically, hippocampal neurons appear to retrieve both the appropriate and inappropriate memory representation when mPFC is inactivated (Navawongse and Eichenbaum, 2013), indicating that mPFC is not necessary for memory recall per se, but that it is required to select the appropriate memory given the context.

Another line of research providing evidence for hippocampal-PFC interactions comes from reports of increased synchronisation of neural activity between hippocampus and prefrontal cortex during replay (Benchenane et al., 2010; Siapas et al., 2005), suggesting a transfer of information between these regions during periods of rest.

Taken together, these studies demonstrate the importance of interactions between hippocampus and (medial) PFC for the integration of new experiences into existing knowledge, memory consolidation and memory expression.

1.4 Thesis Overview

In my DPhil, my aim was to uncover the role of cortical E/I balance and hippocampal-cortical interactions in memory in humans, by using a combination of behavioural measures, interventions that aim to disrupt E/I balance, and neuroimaging.

The work presented in Chapter 4 was motivated by the notion that neuromodulators are thought to regulate plasticity and thus the formation of memories. A potential mechanism for neuromodulators to accomplish this would be by transiently perturbing E/I balance. Therefore, changing the concentration of a neuromodulator during learning would disrupt this process and result in perturbations in the memories formed. To test this idea, I used a pharmacological intervention to increase noradrenaline during associative learning, and measured the effects on neural representations, the concentration of cortical GABA, and on the memories formed. Increased noradrenaline led to a decrease the concentration of cortical GABA, and this

disinhibitory effect predicted neural and behavioural overgeneralisation of associations. These results indicate that inhibition normally facilitates accurate memory formation, and that neuromodulators can tip the balance between stability and plasticity in the brain to favour either depending on the situation.

In Chapter 5, I investigated the neural mechanisms that prevent interference between overlapping memories. Associative memories in neocortex are stored in a dormant state through inhibitory synaptic plasticity following learning, but could these inhibitory connections also play a role in protecting similar memories from interfering with each other? I found that in hippocampus, representations of overlapping memories are separated using contextual cues. To investigate the role of inhibition in neocortical representations, I used tDCS brain stimulation to reduce cortical GABA. Lowering inhibition increased memory interference, supporting the hypothesis that inhibition prevents neocortical memories from coactivating. Furthermore, disturbing E/I balance disrupted hippocampal-neocortical interactions, raising the possibility that in an unperturbed state, hippocampal-neocortical interactions may underlie cortical memory reactivation.

In Chapter 6, I investigated the mechanism that coordinates selective and precise recall in humans. In this study, I applied a novel imaging sequence that combines measures of event-related functional brain activity and real-time changes of neural chemical in the brain. Disinhibition could transiently release memories from the blanket of inhibition, which would imply that memory recall should be accompanied by a decrease in GABA. Indeed, I show that recall of a visual cue from a paired associate is accompanied by a transient decrease in GABA and increase in glutamate in visual cortex. Furthermore, this change is predicted by activity in hippocampus, supporting the hypothesis that hippocampus mediates disinhibition.

Taken together, the results of these experiments reveal how E/I balance, hippocampus, and their interactions contribute to a stable yet flexible memory system.

CHAPTER 2

METHODS FOR INVESTIGATING BRAIN ACTIVITY AND EXCITATORY-INHIBITORY BALANCE IN HUMANS

In order to investigate the neocortical and hippocampal contributions to stable yet flexible memory in humans, I employed a range of techniques to measure and perturb neural activity. To measure activity across the whole brain, I used functional Magnetic Resonance Imaging (fMRI), a non-invasive technique that indirectly measures brain activity as changes in oxygenated blood flow. When combined with approaches like repetition suppression and representational similarity analysis, fMRI can be used to access neural representations across the brain. To obtain a measure of excitatory and inhibitory tone in neocortical areas, I used Magnetic Resonance Spectroscopy (MRS) to measure the concentration of the main excitatory and inhibitory neurotransmitters in the brain. To perturb the balance between excitation and inhibition I used two different approaches: in Chapter 4 I used a pharmacological intervention, and in Chapter 5 I used transcranial Direct Current Stimulation (tDCS), a brain stimulation technique. In this chapter I will review each of the methods mentioned above. I will introduce the general principles of fMRI and MRS, and cover the steps involved in data analyses for both modalities. I will then discuss the neuromodulatory techniques used in this thesis.

.

2.1 Functional MRI

Magnetic Resonance Imaging (MRI) is a non-invasive technique that allows researchers and clinicians to study brain structure and function. Functional MRI (fMRI) is a sub-type of MRI with which brain activity can be assessed. MRI and fMRI have the same underlying biophysical principles, and therefore I will first cover those before moving on to fMRI analysis approaches.

2.1.1 MRI fundamentals

Physics principles

The fundamental measurement principle of MRI is based on a quantum mechanical property called 'spin' exhibited by some atomic nuclei. For the purposes of understanding MRI it is sufficient to interpret spin as rotation rather diving deep into quantum mechanics, which is beyond the scope of this thesis. Due to their spin, positively charged atomic nuclei, including hydrogen (^{1}H, or 'proton') which is of interest in MRI of the brain, have a small magnetic moment. Outside a magnetic field the nuclei are randomly aligned, but when placed in a magnetic field they become either parallel or antiparallel aligned to the direction of the magnetic field. The result is a net magnetisation vector (M) in the direction of the magnetic field (B_0). Here, the nuclei spin or precess at a specific frequency, termed the 'Larmor frequency' ω_o, which is defined as:

$$\omega_o = \gamma * B_0 \tag{2.1}$$

Where B_0 is the magnetic field strength of the MRI scanner in Tesla (T) and γ the nucleus dependent gyro-magnetic ratio in MHz/T.

Besides the B_0 field, the MRI scanner is also equipped with Radiofrequency (RF) coils that generate an additional magnetic field called B_1, oriented perpendicular to B_0. When a B_1 pulse is applied, the magnetisation M will move from aligning with B_0 (the longitudinal plane) to aligning with B_1 (the transverse plane). Following this pulse, protons gradually return into alignment with the longitudinal plane, a process which is called relaxation. The speed with which protons return into alignment with the longitudinal plane depends on their local magnetic environment, i.e. the surrounding tissue. The same RF coil used for flipping the protons into the transverse plane also acts as a receiver coil: the protons spinning back into alignment with the B_0 field induce a decaying current which is picked up by the RF coil.

Localisation

The final element to discuss for a basic understanding of MRI is localisation. If all protons were to experience the same B_0 magnetic field, they would all precess at the same frequency in accordance with Equation 2.1. To enable identification of where in the brain the measured proton signal is coming from, the magnetic field strength is spatially varied using gradient coils.

The gradient fields lead to protons experiencing different magnetic fields depending on their location in space, and thus to have different resonant frequencies. Subsequently, by applying RF pulses at different frequencies, protons in different parts of the brain can be targeted. Protons are typically excited in slices along the longitudinal plane. The resulting signal picked up by the RF coil will be composed of many different frequencies that act as a spatial 'tag' indicating where the signal came from. The measurements by the RF coil in the frequency domain, called k-space in the context of MRI, are transformed into the spatial domain through a Fourier transform.

MRI contrasts

After the RF coil is turned off, the time it takes for the magnetisation M to return to the longitudinal plane is characterised by the time constant T1. T1 differs per tissue: e.g. ^{1}H atoms in fat tissue will return to the B_0 plane fast while ^{1}H atoms in water will take longer. The decay of magnetisation in the transverse plane is characterised by the time constant T2. T2 is not only affected by protons spinning out of the transverse plane, but also by the dephasing of proton spins. Nuclei interact with neighbouring nuclei, and these interactions cause spins to become desynchronised. In addition, the dephasing of spins is also affected by local field inhomogeneities. The time constant T2*, or effective T2, incorporates both sources of dephasing. Critically, T1 and T2/T2* relaxation are simultaneous but independent processes. By optimising an MRI sequence for sensitivity to T1, T2, or T2* relaxation time, the amount of signal obtained from different tissue types can be tuned. The main parameters that are typically tuned are Echo Time (TE) and Repetition Time (TR). TE refers to the time between application of an RF pulse and recording of the signal by the RF coil, and TR refers to the time between two RF excitation pulses. The longer the TR, the more time there is for magnetisation to recover completely. T1-weighted images are often used for structural MRI and are best acquired with short TE and short TR, as this maximises the difference between white matter, grey matter and cerebrospinal fluid (CSF). T2*-weighted images are used for fMRI and use a longer TE which is tuned to the T2 of grey matter.

2.1.2 The BOLD signal

MRI can be applied to measure functional brain activity in response to a stimulus or at rest, which is aptly termed functional MRI (fMRI). fMRI relies on the Blood Oxygen-Level Dependent (BOLD) signal: when a neuronal population becomes active (i.e. release action potentials), more oxygen is delivered to this region through blood vessels running through the brain to fuel the neural activity. Oxygenated blood and de-oxygenated blood have different magnetic properties which particularly affect T2* relation time, and therefore it is possible measure the effect of an influx of oxygen-rich blood. For this reason, fMRI employs sequences sensitive to the T2* contrast.

It is important to note that, by relying on changes in blood flow, fMRI is an indirect measure of brain activity. This has consequences in the interpretation of fMRI data. The relationship between neural activity and changes in blood flow is termed neurovascular coupling, and this is not a fast process: local changes in blood flow are relatively slow (in the order of seconds) compared to neural activity which occurs in the order of milliseconds. Blood flow changes are characterised by the Haemodynamic Response Function (HRF) (Figure 2.1). The characteristic shape of the HRF arises as follows: immediately following neural activity there is a small delay or even decrease in BOLD due to consumption of oxygen. In response, more oxygenated blood is directed to the affected area and the actual oxygen demand is exceeded, leading to an excess of oxygenated blood. This phase peaks ~5-10 seconds after onset of neural activation (Logothetis, 2003). Oxygen subsequently undershoots and only fully returns to baseline ~30 seconds after the neural event. The implication for fMRI is that temporal resolution, which is already limited by TR, is effectively reduced even further: all neural activity is passed through a filter that turns a millisecond process into a 10+ second one.

Another complication relevant to the interpretation of BOLD-derived measures of neural activity is that not only excitatory neurons contribute to the BOLD signal. Studies in rodents show that other neuronal cell types, like astrocytes and certain types of interneurons also contribute to fMRI BOLD signal (Howarth et al., 2021; Lee et al., 2020; Takata et al., 2018). Furthermore, the neurons that are responsible for the majority of blood oxygen consumption appear to be different from neurons the that regulate increases and decreases in blood flow (Howarth et al., 2021). These caveats should be kept in mind when interpreting changes in BOLD-derived neural activity.

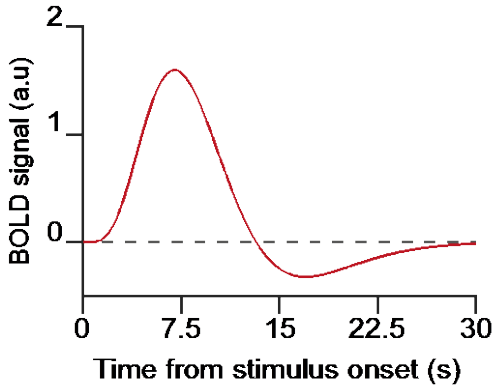


Figure 2.1 The haemodynamic response function. "a.u." refers to "arbitrary units".

EPI and Multiband

All fMRI data in this thesis was collected using Echo Planar Imaging (EPI) sequences. EPI is a rapid imaging technique where all slices in each brain volume are collected within a single excitation. This is achieved by very rapidly switching Gx and Gy gradients to scan the entirety of k-space. The benefit of using EPI sequences is a greatly reduced scan duration. Disadvantages are that a longer read-out time is necessary to collect an entire volume, and during this time T2* signal will have decayed, leading to more blurring and thus lower spatial resolution.

To counteract the loss in spatial resolution, EPI can be used together with acceleration methods such as Multiband acceleration. Using multiband techniques, multiple slices can be acquired simultaneously. The data of multiple slices is collected at the same time and reconstruction of the slices is achieved by taking into account the distance of the slices to the RF coils: the signal of a nearby slice will be stronger than from a slice that is further away. Multiband acceleration factor refers to the number of slices collected at the same time, e.g. multiband factor 2 means 2 slices are acquired simultaneously. The increase in acquisition speed comes with an SNR penalty and can generate artefacts, making multiband factors beyond 4 less common.

2.1.3 fMRI preprocessing

In this section I will briefly introduce the standard analysis pipeline for an fMRI dataset. All fMRI data processing and analyses in this thesis were performed in Statistical Parametric Mapping (SPM) (https://www.fil.ion.ucl.ac.uk/spm/), and therefore I will adhere to SPM's terminology for the different processing steps.

The purpose of fMRI data analysis is generally to identify which brain regions respond significantly to a particular stimulus across different individuals, which is achieved by generating a group-average activity map of the brain for that stimulus of interest. As all people's brains have

different shapes and sizes, their brains need to be somehow aligned. Furthermore, different sources of noise need to be accounted for and where possible removed from the data. To determine which brain regions respond significantly to the stimulus of interest, either a univariate or multivariate approach can be used. In a univariate analysis, a model is fitted to each individual voxel, whilst in a multivariate analysis a model is fitted to groups of voxels. Further statistical analysis is then performed on the parameters fitted by the model. In the subsections below I will cover these processing steps in more detail.

Slice timing correction

In an EPI sequence, all brain slices are not acquired at the same moment in time. This means that there will be systematic timing differences between the different slices in an fMRI volume. fMRI sequences differ in the order in which the slices are acquired (sequentially or interleaved) but regardless of this, the timing between the different slices needs to be matched before further processing. This can be achieved through slice timing correction, where a single reference timepoint is chosen for each fMRI volume and the signal in each slice is shifted and interpolated such that it matches the reference timepoint.

Bias and distortion correction

Some artefacts in fMRI data can be corrected during preprocessing. The B_1 field of the MRI scanner can induce local RF inhomogeneities, degrading the quality of the fMRI data and leading to low frequency intensity variations across the images. One of the first preprocessing steps is to correct for this effect using bias correction.

fMRI is particularly susceptible to B_0 magnetic field inhomogeneities which cause distortion and 'drop-out' in the fMRI images. To correct for these types of artefacts, a set of 'fieldmap' images is usually collected for each study participant during their MRI scan. The fieldmap images characterise the magnetic field in the scanner as experienced by the brain located in it. Two images are collected in such a way that they experience distortions in opposite directions, and their midpoint should approximate an undistorted image. The transformation from these fieldmap images to the midpoint between them is then used to infer how much every voxel in the fMRI images should be shifted to correct for magnetic field inhomogeneities.

Motion correction

fMRI is typically employed to localise the brain region(s) where a certain process or computation occurs. Therefore, it is important that the source of the signal that is measured can be reliably identified. Voxel sizes are typically in the order of millimetres, and it reasonable to expect a study subject will move at least a few millimetres over the course of a long research scan. For this reason, motion correction is a very important processing step, as otherwise a voxel will not correspond to the same position across the entire acquisition.

In a typical fMRI analysis pipeline, motion is usually corrected for at two stages: during preprocessing and during statistical modelling. During preprocessing, all the volumes in an acquisition are aligned to a reference volume, typically the first volume. Linear motion realignment algorithms treat the brain as a rigid body and translate and rotate each volume on the x, y and z-axes to minimise the difference with the reference volume. Non-linear methods that include warping manipulations also exist, but only linear methods were used for the work in this thesis. The estimated translation and rotation parameters for each volume are saved for additional motion correction: they are included in the statistical modelling as nuisance regressors.

Coregistration and Normalisation

To perform a group-level analysis, fMRI acquisitions of different study subjects need to be averaged. Every person's brain anatomy is unique, and every person's brain will be positioned in a different position in the MRI scanner. Therefore, acquisitions from different participants need to be aligned to each other, which is normally achieved through normalisation to a template brain image. A commonly used template is the MNI152 template. This is an averaged T1-weighted structural scan of 152 healthy individuals from the Montreal Neurological Institute (MNI). Normalisation to the MNI template is typically performed in multiple steps and requires both the fMRI dataset as well as a structural T1 scan for every participant. The structural scan is first normalised to the MNI template: due to the higher spatial resolution of the structural image, this normalisation is more precise than if the functional images were used. Then, the functional images are coregistered to the non-normalised structural image, and finally the coregistered functional images are normalised to the MNI template, using the saved transformation parameters from the structural-to-MNI normalisation step.

Smoothing

For univariate analyses, fMRI volumes are typically smoothed using a 3D Gaussian kernel with a full-width at half-maximum (FWHM) of 5-8 mm depending on the voxel size of the fMRI data. Smoothing averages out random noise and also has the benefit of further reducing anatomical differences between different participants' scans. This step is skipped for multivariate analyses which take advantage of intensity variation along neighbouring voxels.

2.1.4 fMRI univariate statistical analysis

To identify regions in the brain that respond to the stimulus of interest, the typical univariate approach is to take the time course of the BOLD signal in each voxel and regress onto this the time course of the stimuli or neural process of interest (e.g. at which time points a stimulus was presented). To this end, the General Linear Model (GLM) is used:

$$Y = \beta X + \varepsilon \tag{2.2}$$

The GLM is a subtype of linear regression, where the BOLD signal time courses of every voxel represent the response variables we want to predict, Y, using a design matrix, X, which contains the experimental variables of interest. The weights (or relative contributions) of each variable in X, β, are the parameters that are optimised to minimise the error between the data and the model. The parameter ε represents the error of the model which is minimised.

The design matrix X contains all the experimental variables of interest. It includes the timepoints where stimuli were presented or where the participant completed an action of interest (e.g. pressing a button), as well as nuisance regressors: e.g. the motion estimates from the motion correction preprocessing step. Any physiological measurements like pulse and breathing can also be included here. The regressors for the stimuli of interest in the design matrix are generated as follows: at each point in time, stimulus onsets are modelled either as present (1) or not (0), optionally modulated by another parametric regressor (if the stimulus had several discrete levels). For designs where stimuli are presented in a blocked manner, this results in a boxcar signal, and in case of an event-related design, where stimuli are not presented in blocks, this results in delta impulses. This time course is then convolved with the hemodynamic response function (Figure 2.1).

The optimisation of the β parameters in Equation 2.1 is achieved by using Ordinary Least Squares (OLS):

$$\beta = (X^T Y)^{-1} X^T Y \quad (2.2)$$

Other processing steps usually included at this stage are a correction for autocorrelation and high-pass filtering. Autocorrelation is present in the fMRI images as the signal in a voxel will be correlated to signal in nearby voxels and therefore their time courses will not be independent. A high-pass filter is applied to the data to account for scanner drift and physiological noise which are also sources of autocorrelation across voxels.

The result of these operations is a β map for each regressor included in the model. For each study participant, beta maps corresponding to the stimulus of interest are usually contrasted with beta maps corresponding to a control condition. By looking at the difference between conditions, the voxels where activation is specific to the condition of interest can be identified. At this point, the contrast images of individual study participants are usually combined in a group or second-level analysis. A group average contrast image is generated, and a T- or F-test can then be performed to test which voxels had a significant response to the stimulus of interest. As the significance test is performed on many voxels individually, any result needs to be corrected for multiple comparisons to avoid false positives.

2.1.5 fMRI multivariate statistical analysis

Using univariate fMRI analyses, all voxels are analysed individually and independently. However, it is unlikely that the brain only processes information in independent voxel-sized blocks. Therefore, multivariate approaches take into account activity across groups of voxels. Presentation of a stimulus is thought to induce a pattern of activity across neighbouring voxels (Haxby et al., 2001). Indeed, stimuli can be differentiated from each other using their evoked patterns of activity in a brain region (Cox and Savoy, 2003). The purpose of multivariate techniques is not to replace univariate analyses in fMRI, but rather to extend analyses and make different claims about the neural processes at play. A wide variety of multivariate techniques have been developed or adapted for use with fMRI data, including Principal Component Analysis (PCA), Support Vector Machines (SVMs) and Multi-Voxel Pattern Analysis (MVPA). Here, I will focus on MVPA, as I applied an MVPA-based approach in Chapter 5.

With MVPA, groups of voxels are jointly analysed, either in a Region of Interest (ROI) or using a searchlight approach across the entire brain. MVPA relies on the presence of neural structure

across voxels within a brain region. For instance, in the visual cortex neurons are clustered according to the orientation of a stimulus they respond to (Bartfeld and Grinvald, 1992; Obermayer and Blasdel, 1993). The presence of neuronal clusters would induce significant activation changes within a voxel which should be detectable using fMRI. Indeed, MVPA has been successfully applied in fMRI to decode stimulus orientation (Kamitani and Tong, 2005), as well as more high-level information like object categories (Haxby et al., 2001).

Representational Similarity Analysis

A widely used variant of MVPA is pattern-similarity MVPA, commonly known as Representational Similarity Analysis (RSA) (Kriegeskorte et al., 2008a). RSA is typically used in event-related experiments rich in conditions or stimuli. Instead of analyzing patterns of voxels directly, in RSA the differences between voxel patterns are evaluated (Diedrichsen and Kriegeskorte, 2017; Kriegeskorte et al., 2008a).

The analysis approach is as follows (Figure 2.2): first, for every trial, the pattern of activity within a (searchlight) ROI is extracted from the preprocessed data. Each pattern is then labelled according to the trial condition, and subsequently the pattern of each trial is correlated with the patterns extracted for all other trials. The correlations are then summarised using a correlation distance measure, such as Euclidean distance. The result is a collection of correlational distances, also called dissimilarities, for each trial-to-trial comparison within the experiment. RSA thus moves from a multidimensional voxel space into a stimulus similarity/dissimilarity space. Typically, the dissimilarities are presented in a matrix format, termed Representational Dissimilarity Matrix (RDM). To test hypotheses relating to what information might be stored within the neural activity patterns, the RDM resulting from the trial-to-trial correlational distances is compared to a model RDM. The model RDM can be based on stimulus features, behaviour or a computational model. In addition, researchers have used RSA to compare neural data across different modalities and species: Kriegeskorte and colleagues showed that a set a of visual stimuli give rise to equivalent RDMs in the human Inferior Temporal (IT) cortex measured using fMRI, and macaque IT cortex measures using electrophysiology (Kriegeskorte et al., 2008b). This study illustrates a key strength of RSA over univariate methods: moving away from voxel space to dissimilarity space allows neural representations to be measured and compared across datasets collected using different experimental modalities.

RSA has been successfully applied in fMRI to study a wide range of neural processes, including visual processing (Hiramatsu et al., 2011), motor representations (Wiestler et al., 2011), and

memory (Xue et al., 2010). In Chapter 5, I apply RSA to study context-dependent relational codes in hippocampus.

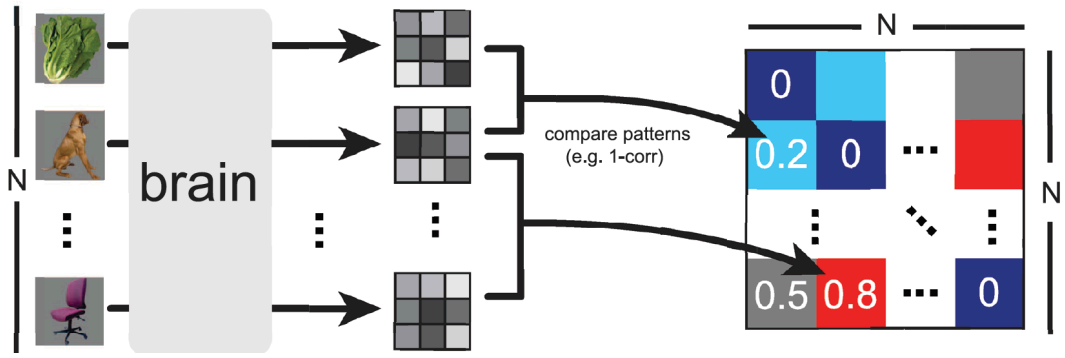


Figure 2.2 Illustration of representational similarity analysis. In this example, the presentation of N visual stimuli results in an activity pattern across voxels for each stimulus. The activity patterns are compared to each other and the correlational measures are entered into a matrix. Adapted from (Nili et al., 2014).

2.1.4 Repetition suppression

A shortcoming of using fMRI to study neural representations is its limited spatial resolution. While neurons may be clustered in subpopulations that activate as a population (Bartfeld and Grinvald, 1992; Obermayer and Blasdel, 1993), a 3 mm^3 voxel containing in ~10^5 neurons will contain a multitude of clusters making it more challenging to locate the subpopulations involved in a particular neural computation. One solution to this problem is to use multivariate analysis methods. However, univariate fMRI can also be used to access representational information in the brain, by relying on the phenomenon that neurons exhibit a relative suppression in their activity in response to repeated presentation of a stimulus they respond to. This phenomenon is referred to as "repetition suppression" or "fMRI adaptation" and has been observed across a variety of brain regions, stimulus types and timescales (Baylis and Rolls, 1987; Miller et al., 1991; Riches et al., 1991).

While repetition suppression paradigms are typically used to access representations of individual stimuli (Grill-Spector et al., 2006), they can also be applied to index representational overlap between different stimuli, termed Cross-Stimulus Suppression (XSS) (Barron et al., 2016b). Cellular evidence for this phenomenon comes from single-unit recordings in the macaque IT; Sawamura and colleagues showed that certain neurons responding to stimulus A and B individually both show suppressed activity not only for A preceded by A (repetition suppression), but also for A preceded by B (XSS) (Sawamura et al., 2006). Thus, XSS can be used to measure representational overlap and has been applied to index associative memories (Barron et al., 2016a) and the neural overlap between a stimulus and reward (Boorman et al., 2016).

Repetition suppression and XSS allow for the non-invasive measurement of neural computations in humans that previously only could be achieved using animal models, for instance to study grid-like codes in hippocampus (Constantinescu et al., 2016; Doeller et al., 2010). As our understanding of complex cognitive processes depends mainly on studies in humans, fMRI adaptation paradigms are a promising tool to gain mechanistic insight into human cognition. In Chapters 4 and 5, I applied XSS paradigms to study the overlap between contextually overlapping neural representations. In the next section, I will discuss the models that have been proposed to explain the origin of the repetition suppression phenomenon.

Biophysical model suggested to underly repetition suppression

Repetition suppression observed at the single-cell level has been reported in various experimental conditions, e.g. with various interstimulus intervals and with the inclusion of other interleaved stimuli, and in various brain regions, including IT cortex, Prefrontal Cortex (PFC), Primary Visual Cortex (V1) and entorhinal and perirhinal cortex (Albrecht et al., 1984; Baylis and Rolls, 1987; Blakemore and Campbell, 1969; Li et al., 1993; Miller et al., 1991; Riches et al., 1991). Repetition suppression thus seems to be a general property of neurons acting as a short-term memory mechanism.

Synaptic fatigue has been proposed as an underlying mechanism for repetition suppression. Synaptic fatigue could either emerge as a result of reduced action potential firing due to hyperpolarisation following a previous action potential, or as a result of a synaptic depression mechanism following an action potential, reducing responsivity to synaptic inputs following activity (Abbott et al., 1997; Carandini and Ferster, 1997; Sanchez-Vives and McCormick, 2000). This model predicts that neurons that respond most strongly to a stimulus show the largest decrease in activity when the stimulus is repeated.

An alternative explanation for repetition suppression is neural sharpening. The sharpening model posits that with repetition of a stimulus, information might be encoded more efficiently or 'sharper', by activating fewer neurons (Wiggs and Martin, 1998). This model suggests that not only the amplitude of the response to the second stimulus changes, but also the neural pattern of activation. This combination of change in amplitude and pattern of activation has been observed in response to presentation of familiar stimuli, but not for repeated presentation of novel stimuli. However, this model is consistent with the experimental observation that neurons most tuned to a particular stimulus feature exhibit less repetition suppression, which is opposite of what the fatigue model would predict (Baker et al., 2002; Freedman et al., 2006).

Finally, a third model providing an account for repetition suppression is the facilitation model, which accounts for behavioural priming effects. The facilitation model posits that with repeated presentation of a stimulus, the speed at which it is processed increases (Grill-Spector et al., 2006). Mechanistically, this could be accounted for by shortened response latencies or firing durations due to expectation priming, which is consistent with the predictive coding account for brain activity. The predictive coding framework states that higher cognitive areas top-down modulate activity in sensory areas by predicting upcoming inputs (Friston, 2005). Thus, repetition suppression could result from top-down modulation. There is evidence from an fMRI study suggesting this model is plausible (Summerfield et al., 2008), however there is also evidence from a magnetoencephalography (MEG) study suggesting that repetition suppression and expectation effects are two distinct temporal processes (Todorovic and De Lange, 2012).

While no one of these three models offers a complete account for the precise underlying neural mechanism, repetition suppression offers a valuable tool to study cognition, as long as care is taken in the experimental design that expectation effects are controlled for.

Repetition Suppression in fMRI

How does a neural feature observed at the cellular level translate to the BOLD signal? Repetition suppression can be observed using fMRI using appropriate stimulus presentation designs (Figure 2.3). While the results of fMRI in humans show close correspondence with results from single-unit recordings in animal models (Grill-Spector et al., 1999; Sawamura et al., 2006), it would be simplistic to assume the exact same process is at play, given the difference in spatial and temporal resolution, and given that fMRI BOLD is only an indirect measure of neural activity. Indeed, rather than spiking activity, the BOLD signal in general appears to reflect local field potentials (LFPs), i.e. the summation of excitatory and inhibitory potentials in the extracellular space (Logothetis et al., 2001; Maier et al., 2008). The BOLD signal could thus be interpreted as a reflection of the local environment that neurons experience in a voxel. While LFPs are not as straightforward to interpret as neural spikes, most neurons are highly interconnected locally and typically receive their inputs from neighbouring neurons 1-2 mm away (Amir et al., 1993; Douglas and Martin, 2004). This implies that the BOLD signal still reflects local functional activity. However, it is important to remember that the BOLD signal is a measure of blood flow. While blood flow adaptation correlates with neural adaptation effects, the response does not scale (Moradi and Buxton, 2013). The implication of this is that the BOLD signal likely underestimates repetition suppression on the neural level. Despite this, although interpretation of repetition

suppression measured using fMRI is not straightforward, correlation between fMRI repetition suppression and neural repetition suppression suggests it is a valid measure to localise neural suppression effects within the brain.

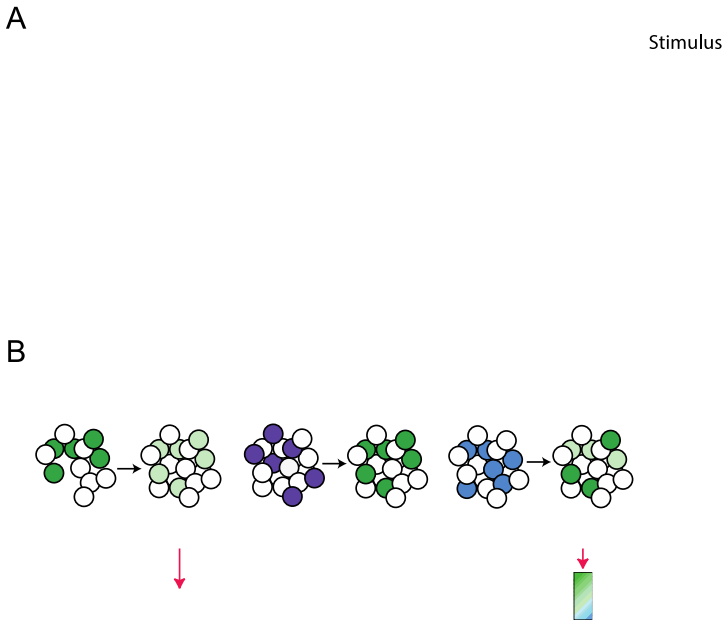


Figure 2.3 Schematic illustrating stimulus presentation design to measure XSS effects using fMRI BOLD. **A** BOLD signal in response to separate presentation of stimuli X, Y and Z is the same, because the average neural activity is the similar for each stimulus. **B** BOLD signal in response to repeated presentation of the same stimulus (X-X), a stimulus preceded by a non-overlapping other stimulus (Y-X) and a stimulus preceded by an overlapping stimulus (Z-X). Adapted from (Barron et al., 2016b)

Comparing Repetition Suppression with multivariate methods

Both repetition suppression and multivariate methods can be used to access sub-voxel representations, but while multivariate methods evaluate representations across voxels, repetition suppression only probes within-voxel representations. Thus, as these two categories of methods evaluate different aspects of fMRI data, they can be used to explore different questions. For instance, using repetition suppression a specific biophysical phenomenon can be probed, while with RSA neural representations can be studied where no specific neural mechanism is targeted. There is, however, evidence suggesting that the two methods are very highly correlated (Sapountzis et al., 2010).

2.2 MR Spectroscopy

Proton-Magnetic Resonance Spectroscopy (^{1}H-MRS, or simply MRS) is a non-invasive technique used to measure the concentration of different neurochemicals and can, therefore, be used to investigate the underlying physiology of the brain. The technique leverages the same physical properties of protons as regular MRI to obtain a measurement and can be acquired with the same equipment. In this section I will discuss the underlying physical principles relevant to MRS, as well as the data analysis steps involved in typical single voxel MRS.

2.2.1 MRS physics principles

As mentioned in Chapter 2.1.1, protons have a magnetic moment called spin that results in resonance when they are subjected to a magnetic field. A proton's resonant frequency depends on its Larmor frequency and the strength of the magnetic field it experiences (Eq. 2.1). When subjected to a RF pulse, protons will align with the RF field and precess at their characteristic frequency. The key physical property of protons that allows for MRS measurements is that the precise resonant frequency of a proton depends on the molecule they are bound to; the resonant frequency of a proton is affected by the other molecules around it (their chemical and electrical environment), referred to as chemical shift, as well as the configuration of electrons along the different atoms of a molecule, called J-coupling. These properties are leveraged in MRS. MRS can also be used to quantify other nuclei besides protons, e.g. ^{13}C or ^{31}P (Befroy and Shulman, 2011; Rothman et al., 2011). However, ^{1}H MRS is most commonly used as hydrogen is part of many molecules of interest in the brain. Furthermore, no additional hardware is needed to perform this type of measurement with a standard MRI scanner.

Chemical shift and J-coupling

Chemical shift is the main molecular property through which an MRS spectrum is generated. It refers to the phenomenon that a proton's spin is not only determined by its Larmor frequency and the magnetic field, but also by the atoms around it. Surrounding atoms shield a proton from the magnetic field, and thus the magnetic field experienced by a proton is dependent on the molecule it is part of.

The other physical phenomenon that affects a proton's resonant frequency is J-coupling. Whereas chemical shift only applies to the nuclei bound directly to the proton, J-coupling encompasses the interactions of the proton with any nuclei that are part of the same molecule

through chemical binding (e.g. through sharing of electrons). This phenomenon effectively splits up the resonance of a molecule in several characteristic multiplet structures.

The combination of chemical shift and J-coupling effects give each molecule a characteristic resonance profile or 'spectral fingerprint' which can be detected using MR. Importantly, chemical shift scales with the strength of the magnet field. Therefore, at higher field strengths proton resonances shift more and become more separated, which makes it easier to determine metabolite concentrations. Thus, high-field MRS (≥3 T) has risen to prominence. Furthermore, unlike chemical shift, J-coupling is not affected by magnetic field strength. This means that higher field strengths are even more advantageous as the resonance effects due to chemical shift and J-coupling are more easily distinguished and molecules become more easily identifiable.

The MRS spectrum

Similarly to MRI, in MRS protons are excited using an RF-pulse and their resonance is measured using the RF coil. In MRS, the signal measured from the RF coil is then moved from the time-domain to the frequency-domain by applying a Fourier transform, resulting in a characteristic MRS spectrum (Figure 2.4). Each peak in this spectrum represents a group of 1H nuclei that are part of a particular molecule. Given a magnetic field strength, the frequency of this peak will always the same for that molecule. Furthermore, the size of the peak is directly proportional to the concentration of that molecule in the measured area. MRS spectra in Hertz (Hz) are normally converted to a normalised unit of measure using a reference frequency, typically of N-acetylasperate (NAA) which is set at 2.02 parts per million (ppm). This aids the comparison of spectra acquired at different magnetic field strengths.

The main methodological challenge of MRS is that the concentrations of molecules of interest are very low, particularly in comparison to water: water is present in the molar range while e.g. neuromodulators like dopamine are present in the micromolar range. While it is currently not possible to measure concentrations at the level of micromoles, it is possible to measure concentrations of the molecules in millimolar range. However, to allow for sufficient signal, MRS is subject to several restrictions. First, MRS acquisitions are typically only performed using one large voxel, e.g. 2 cm^3, as the signal strength scales with the number of metabolites in the volume of interest (VOI). Second, acquisitions are typically over a longer period of time, and the measurements are averaged across this time period to increase in signal-to-noise ratio (SNR). To obtain a single measurement of metabolite concentrations, 128-256 measurements are typically averaged, which takes 4-8 minutes. There is an active effort to expand the spatial and temporal

limits of MRS, and recently both multivoxel MRS (MRS Imaging, MRSI) and time-resolved functional MRS (fMRS, Chapter 3) have become more prominent.

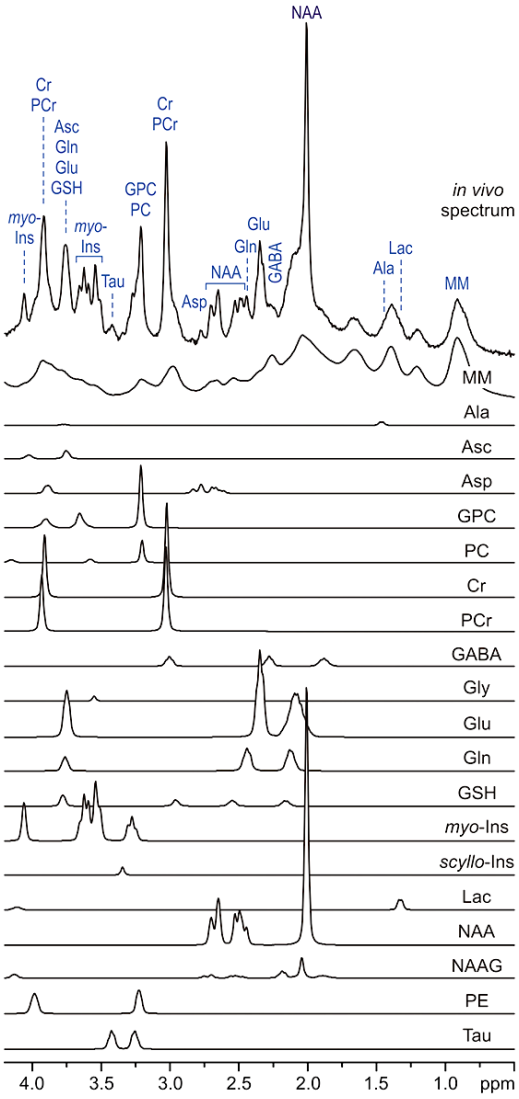


Figure 2.4 Example 7T in vivo MRS spectrum and fitted spectra for 20 metabolites and macromolecule signal (MM). Adapted from (Godlewska et al., 2017).

2.2.2 The MRS acquisition

Shimming

If the magnetic field varies over the VOI (typically a 2 x 2 x 2 cm^3 voxel), resonant frequencies of the same molecule will vary throughout the voxel. Variable resonances will lead to broader peaks in the resulting spectrum and more overlap between peaks, resulting in less precise metabolite quantification.

There are numerous sources causing magnetic field inhomogeneities, including placement of a head inside the scanner bore. To minimise these distortions in the magnetic field, a correction

technique called shimming is used before any MRS acquisition is made. Using gradients, the magnetic field is adjusted locally to be as homogeneous as possible. Once shimming has been performed, it is important that the head stays in the same position, otherwise the corrections in the magnetic field will no longer be accurate.

Water Suppression

Water is the most abundant molecule in the brain and is present at a concentration ~1000 times higher than metabolites of interest. If left unsuppressed, the resonance frequency of water protons will overlap with the resonance frequencies of the metabolites of interest and distort the entire spectrum's baseline signal. Therefore, several water suppression techniques have been developed. Typically, CHEmical Shift Selective (CHESS) RF pulses are used, which leverage the difference in the chemical shift of water compared to the metabolites of interest (Haase et al., 1985). Using this sequence, water protons are first selectively excited or 'presaturated', after which the signal is 'spoiled' by bringing the excited proton spins out of phase with each other. This process of excitation and spoiling is repeated several times in a row to improve suppression of the water signal before the actual MRS acquisition takes place. For ultra-high field MRS water suppression techniques have been developed that are variants of CHESS, such as the VAriable Power and Optimised Relaxation delays (VAPOR) method (Tkáč et al., 1999). With this sequence, 7 presaturation pulses with optimised flip angles and timings are applied.

The effectiveness of the water suppression is heavily affected by the homogeneity of the magnet field and subject motion. In practise it is difficult to successfully suppress all water signal over the course of an acquisition, and a general rule of thumb is that the residual water signal (the height of the water peak, at 4.65 ppm) should not exceed the strongest signal (tallest peak) in the rest of the spectrum.

An alternative to this presaturation approach of water suppression is to use the MEscher-GArwood (MEGA) selective refocussing method. Briefly, using this method, all proton spins are first excited using a non-selective RF pulse (Mescher et al., 1998). The water signal is then selectively inverted and spoiled by gradient pulses, which do not affect the signal of interest. A single read-out is taken after this, and the process is repeated over each TR.

Eddy current measurement

MRS involves rapid switching of magnetic field gradients, which can induce currents in the area targeted by the gradients. These eddy currents in turn lead to the generation of small magnetic fields which cause a phase modulation of the acquired MRS signal. These phase modulations are problematic as they can mask metabolite signals of interest. Luckily, this issue can normally be easily resolved by acquiring the unsuppressed water signal from the VOI. From this very brief, extra acquisition, the eddy current phase modulation can be deduced and can be applied in reverse to the acquired spectrum. Eddy current correction is performed after the scan using this extra measurement and is usually one of the first steps of preprocessing.

J-difference edited localisation methods

The resonance peaks of metabolites present in low concentrations can be hard to detect or masked by other metabolites, particularly at lower magnet field strengths where resonance frequencies are less spaced out (3 T or less). For this reason, J-difference edited sequences are used in which a specific metabolite can be targeted for quantification. Edited sequences are frequently used to quantify GABA ('GABA-edited spectroscopy'). GABA is only present in low concentrations in the brain and has resonance peaks that are nearby or overlapping with resonances of other more abundant metabolites like NAA or glutamate (Figure 2.4), making it challenging to quantify. With edited spectroscopy, the issue of overlap is addressed. Here, two acquisitions are made for every measurement: the first acquisition is a regular one where a typical spectrum is obtained, while for the second one an extra pulse is added that targets (or 'refocuses') an individual J-coupled resonance and inverts it. The rest of the spectrum of the second acquisition is unaffected. During data processing, the difference between the pair of spectra is calculated, and all signal cancels each other out apart from the targeted J-coupled resonance (Figure 2.5). Edited spectroscopy can be performed with any metabolite that exhibits J-coupled resonance. The benefit of this type of acquisition is the high accuracy with which the concentration of the targeted metabolite can be measured. Downsides include the loss of information on all other metabolites (though information can be obtained from the 'off' spectra), and a longer acquisition time, since two measurements need to be taken for each difference spectrum.

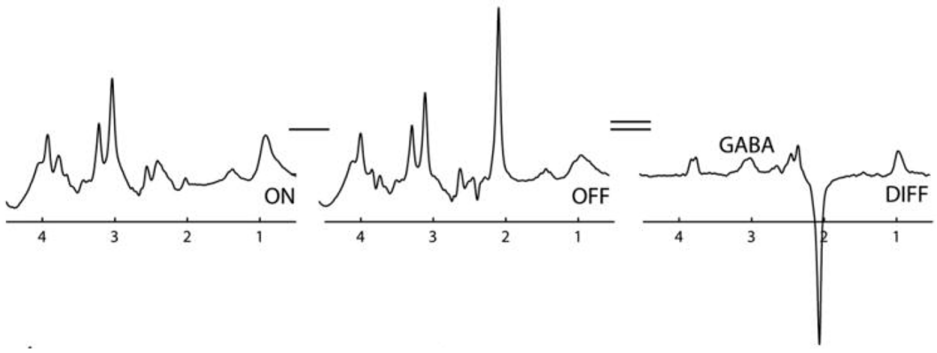


Figure 2.5 Illustration of GABA-edited MRS acquisition. Left: GABA-edited acquisition. Middle: regular acquisition. Right: Difference spectrum. Adapted from (Mullins et al., 2014).

2.2.3 MRS Analysis

Preprocessing of MRS data involves a set of standardised steps that are reasonably comparable across different MRS processing toolboxes. MRS data is acquired in the time domain as Free Induction Decays (FIDs), and all preprocessing is also performed in the time domain. Subsequent model fitting and analysis of the results is performed in the frequency domain.

Preprocessing

During preprocessing of MRS data, several sources of noise are addressed and remedied. The first step in MRS preprocessing is typically eddy current correction using the unsuppressed water signal (Chapter 2.2.2). Next, frequency drift and phase correction are usually addressed. Due to the magnetic field not being completely stable over the duration of an MRS acquisition, and due to respiration and small body movements, small frequency and phase fluctuations are present in the acquired FIDs. Frequency drifts can be corrected by aligning the spectra using methods such as cross-correlation, or by using the characteristic NAA metabolite peak as a reference (Figure 2.4). Phase correction involves the separation of real and imaginary signal contained within the FID. In MRS, we are interested in the real part of the signal and the phase correction ensures that the Fourier transform of the real part of the signal is well aligned.

Despite the inclusion of water signal suppression in the MRS acquisition, some water signal may still remain. Furthermore, the water signal frequency can change over the acquisition (for instance due to subject movement) rendering the water suppression suboptimal. Residual water signal can be removed post-scan, using a black-box method like Hankel-Lanczos Singular Value Decomposition (HSLVD) (Cabanes et al., 2001). Using this method, the peaks of the nuisance signal (in this case water) are modelled and subtracted from the FID.

Metabolite quantification

The objective of MRS analysis is to disentangle the combined spectrum (Figure 2.4) and figure out each molecule's contribution and thus concentration. The key characteristic that allows for quantification is that the height of the peaks is directly proportional to the concentration of the nucleus to which the peak belongs. However, quantification is complicated by overlap between resonance peaks of different nuclei. Quantification methods essentially aim to reproduce the measured spectrum by summing different quantities of each metabolite thought to be in the spectrum. This type of analysis requires previous knowledge on the metabolites thought to be present in the sample. This information is defined in a so-called 'basis set' of metabolites, which is MRI scanner specific and usually determined through controlled experiments or simulations. For each metabolite in the basis set, the model fitting procedure yields a concentration estimate and a goodness of fit parameter called the Cramer-Rao Lower Bound (CRLB), which represents the standard deviation of the concentration parameter estimate. CRLB is one of the parameters evaluated to determine the quality of the data acquisition. Other common measures include SNR and spectral linewidth, defined as the full-width at half-maximum peak height (FWHM) of a reference peak, e.g. NAA. However, it should be noted that evaluating the quality of MRS spectra is not straightforward, illustrated by a study in which spectroscopy experts were asked what defines a good spectrum: the most common response was 'it depends' (Kreis, 2004). Fortunately, efforts are being made by the MRS community to move towards more standardisation and transparency, for instance through the Minimum Reporting Standards in MRS recommendations (Lin et al., 2021).

2.3 Perturbing E/I balance

For the experimental work presented in this thesis, I have used two techniques for manipulating E/I balance: transcranial Direct Current Stimulation and a pharmacological intervention using the drug atomoxetine. I will introduce these two approaches in this final section of this chapter.

2.3.1 Transcranial Direct Current Stimulation

Brain stimulation is a non-invasive method where typically either an electrical current or magnetic pulse is delivered to the brain. In Chapter 5, I used a common type of electrical brain stimulation, transcranial Direct Current Stimulation (tDCS), to modulate cortical GABA.

Modern tDCS has been applied for about 20 years and there is now a sizeable literature documenting its effects and mechanism of action. Anodal tDCS has been applied to enhance cortical excitability, i.e. changes in local neuronal firing rates measured using electrophysiology (Bindman et al., 1962) and modulate GABA-mediated inhibition, measured using paired-pulse Transcranial Magnetic Stimulation (ppTMS) (Nitsche et al., 2005). The effects of tDCS have also been characterised using MRS: changes in neural metabolite concentrations suggest that a decrease in cortical GABA during and immediately following tDCS may be responsible for the increase in cortical excitability (Barron et al., 2016a; Stagg et al., 2009; Stagg and Nitsche, 2011). The duration of the modulatory effects of tDCS depend on the stimulation protocol and can last from minutes to hours after the stimulation has ceased (Bindman et al., 1962; Nitsche et al., 2005). These sustained effects are thought to occur due to a protein synthesis dependent process (Nitsche and Paulus, 2000). tDCS has been applied in various clinical and research settings, including to aid recovery from stroke (Hummel and Cohen, 2006), in the treatment of depression and schizophrenia (Brunelin et al., 2012; Nitsche et al., 2009), and to modulate learning (Jacobson et al., 2012) and decision making (Fecteau et al., 2007).

tDCS is applied by placing two electrodes on the study subject's scalp and running a constant current between them. Current runs from the anodal electrode to the cathodal electrode, and typically the anodal electrode is positioned on the area that is to be stimulated: anodal stimulation has been shown to depolarise neuronal cell bodies, while cathodal stimulation induces hyperpolarisation (Purpura and McMurtry, 1965). However, the assumption that anodal stimulation depolarises while cathodal stimulation hyperpolarises may be an oversimplification, as the effects of cathodal stimulation are particularly inconclusive (Kim et al., 2014; Stagg et al., 2009). To control for placebo effects of the brain stimulation on study subjects' behaviour, oftentimes a 'sham' condition is included in tDCS experiments. Sham stimulation protocols vary, but typically involve ramping up the current and subsequently turning electrical stimulation off.

2.3.2 Pharmacological approaches

Compared to tDCS, pharmacological approaches have the benefit of being able to target a specific pathway or neurotransmitter and can, therefore, be used to reveal the role of a particular neuromodulatory system in a cognitive process. However, pharmacological approaches currently cannot be targeted spatially, as most drugs will have brain-wide effects. This is important to consider when analysing any study results, as the effects measured may not be due to the neuronal

mechanism that was targeted, but instead may be an effect of altered alertness or attention associated with the drug.

To modulate E/I balance pharmacologically, one can either target glutamate or GABA receptors directly, or use drugs that alter neuromodulator concentrations which will have downstream effects on glutamate and/or GABA. Orally administered drugs cannot modulate glutamate, GABA or neuromodulator concentrations directly, but instead change the behaviour of receptors. Agonists increase activity by binding to the receptor similarly to the normal ligand. For instance, glutamatergic agonists are being investigated for the treatment of schizophrenia (Maksymetz et al., 2017), and GABAergic agonists are widely applied in research and for treatment of anxiety and muscle spasms (Krach, 2001; Mondrup and Pedersen, 1984). Antagonists, on the other hand, will decrease activity by blocking the receptor. For instance, glutamatergic antagonists are used in the treatment of Parkinson's disease (Johnson et al., 2009), and the GABA antagonist flumazenil is used as an antidote to overdoses (Hoffman and Warren, 1993). It should be noted that the use of GABA antagonists is difficult in research settings as they have stimulant and convulsant effects. A third mechanism through which drugs can act is by preventing the neurotransmitter from being reabsorbed by the cell, which can be achieved with reuptake inhibitors. Preventing reuptake increases the time where the neurotransmitter will interact with its receptors, and it therefore increases activity of those receptors.

Atomoxetine

In Chapter 4 I make use of the drug atomoxetine, which is a selective noradrenaline reuptake inhibitor. Atomoxetine inhibits the reuptake of both adrenaline (epinephrine) as well as its precursor noradrenaline (norepinephrine) by binding to noradrenaline transporters with high affinity and selectivity (Simpson and Perry, 2003) .

In clinical practise, atomoxetine is prescribed for the treatment of Attention-Deficit Hyperactivity Disorder (ADHD) (Garnock-Jones and Keating, 2009). ADHD is associated with abnormalities in noradrenaline and dopamine as well as reduced metabolism in prefrontal cortex and striatal regions (Katragadda and Schubiner, 2007). Atomoxetine has been shown to increase the concentration of noradrenaline in all brain regions (Koda et al., 2010; Swanson et al., 2006).

Following a single oral dose, atomoxetine is readily absorbed and reaches peak concentration in blood plasma in 1-2 hours. The half-life of the drug depends on whether individuals are extensive or poor metabolisers – the majority of the population (90%) are estimated to be extensive metabolisers where half-life is 5.2 hours (Eli Lilly and Co., 2003). Owing to a gene variation,

atomoxetine's half-life in poor metabolisers is 21.6 hours (Eli Lilly and Co., 2003; Sauer et al., 2005). Therefore, the possibility of some individuals being slow metabolisers and having longer drug wash-out durations should be taken into consideration in any experiment involving atomoxetine.

Atomoxetine has high affinity to noradrenaline transporters and is thought to have minimal affinity to other major neuromodulator transporters like serotonin and dopamine, as well as minial affinity to other types of receptors and ion channels (Gehlert et al., 1993; Wong et al., 1982). However, a recent investigation found that at clinically relevant doses, atomoxetine significantly occupied serotonin transporters in rhesus monkeys, though the effect of atomoxetine on changes in the extracellular concentration of serotonin were not reported (Ding et al., 2014).

Neurotransmitter concentration changes following atomoxetine have been quantified using microdialysis measurements in rodents, where atomoxetine was found to induce a 3-fold increase in extracellular noradrenaline in PFC (Bymaster et al., 2002). Dopamine concentration also increased 3-fold in PFC, which is unsurprising given noradrenaline transporters have similar affinity for dopamine and noradrenaline (Raiteri et al., 1979) combined with a lack of dopamine transporters in this brain region. Atomoxetine did not alter dopamine concentration in striatum or nucleus accumbens, other regions known to be involved in dopamine signalling. No increase in serotonin was observed in PFC (Bymaster et al., 2002). The consequence of atomoxetine affecting both noradrenaline as well as dopamine levels in the brain is that any effects that the drug may have on neural activity and behaviour may be due to changes in noradrenaline or prefrontal dopamine (and perhaps even serotonin). However, if the effects of atomoxetine are evaluated from the perspective of changes in E/I balance or the concentration of GABA, it is not necessarily relevant whether these effects were mediated by changes in noradrenaline or dopamine. This is the approach I took for the interpretation of the results in Chapter 4.

CHAPTER 3

TIME-RESOLVED FUNCTIONAL MRS: A TOOL TO GAIN MECHANISTIC INSIGHT INTO HUMAN COGNITION?

Proton-Magnetic Resonance Spectroscopy (MRS) is a non-invasive brain imaging technique used to measure the concentration of different neurochemicals in vivo in the brain. MRS data is typically acquired across several minutes before being averaged through time to give a steady-state readout. However, this approach masks the more rapid temporal dynamics of neurochemicals that may underlie neural processes relevant to cognition and behaviour. In this chapter I will discuss how recent advances in MRI allow functional MRS (fMRS) to be used to obtain time-resolved measures of neurochemicals. I will discuss how different task designs may be used to optimise fMRS, and present an analysis consideration particularly related to the quantification of time-resolved GABA, a neurotransmitter which resides at relatively low concentration and is therefore more challenging to measure. Finally, I will consider the appropriate interpretation of the resulting measures. Overall, I will show how time-resolved fMRS provides an important tool for gaining insight into the neural mechanisms that underlie human cognition and behaviour.

3.1 Introduction

Proton-Magnetic Resonance Spectroscopy (^{1}H-MRS or MRS) is a non-invasive imaging technique used to measure the absolute or relative concentration of different neurochemicals in the brain. MRS takes advantage of the fact that the local electronic environment of hydrogen protons varies across different molecules. This leads to variance in the characteristic resonant frequency of protons which gives each molecule a distinct spectral fingerprint. A molecule can be detected with MRS if its concentration is sufficiently high and its spectral fingerprint is sufficiently non-overlapping with other more abundant chemicals. By quantifying the concentration of a given molecule, MRS can be used to investigate the physiology of the brain, including how neural metabolite concentrations differ between patients suffering from neurological disorders and healthy controls (Öz et al., 2014).

MRS provides a relatively direct and rich measure of neural processing, as it allows the measurement of the concentration of ~20 different neural metabolites. Notably, measures of two important neural metabolites can be obtained: glutamate and gamma-aminobutyric acid (GABA). Glutamate is the brain's main excitatory neurotransmitter, while GABA is the main inhibitory neurotransmitter, making MRS a method suitable for investigating in vivo plasticity in the brain.

A major challenge for MRS is that the concentration of water in the brain is around 10,000 times higher than the concentration of neural metabolites of interest (Kreis, 2004). The resonant frequency of protons in water molecules must therefore be suppressed to prevent the water signal dominating the MR spectrum. To detect neural metabolites above background noise, particularly those residing at relatively low concentration, MRS has benefited from the introduction of high field MR systems where the SNR scales with the B_0 magnetic field strength of the MRI scanner. For example, when moving from 1.5T to 7T B_0 field strength, the increase in spectral resolution enables delineation of coupled spin systems between molecules such as glutamate and glutamine, and helps dissociate the spectral fingerprints for glutamate and GABA (Tkáč et al., 2001). At high magnetic fields, the enhanced SNR also improves spatial and temporal resolution, allowing voxel sizes of around 2 cm^3 in single voxel spectroscopy, and application of time-resolved functional MRS (fMRS). Thus, rather than averaging across multiple spectra through time to obtain a static measure of neural metabolite concentrations, fMRS can be used to assess dynamic and task-relevant changes in neural metabolite concentrations. This approach has the potential to capture the temporal dynamics of task-related changes in glutamate, GABA and other metabolites, that can be used to inform our understanding of human behaviour in a manner analogous to task-based fMRI.

3.2 fMRS task design: Blocked versus time-resolved

The majority of fMRS experiments to date have employed a 'blocked' design, where experimental conditions within a task are divided up into blocks that span several minutes. Within each block, spectra are then averaged to obtain static metabolite concentrations that can be mapped onto a particular experimental condition. As observed in fMRI studies, blocked designs are efficient and they do not require an explicit model of the predicted neural response (Mechelli et al., 2003). However, this approach obscures the temporal dynamics of neural metabolites related to physiological processes of interest. Furthermore, blocked designs cannot be used when task trials are classified post-hoc or are dependent on participant performance. Finally, like fMRI (Grill-Spector and Malach, 2001), neural metabolite concentrations may be susceptible to expectation effects, where the neural response induced by a stimulus decreases upon repeated presentation (Apšvalka et al., 2015). Indeed, fMRS studies using a blocked design show that task- or stimulus-induced increases in glutamate are larger or only present in the first experimental block (Ip et al., 2017; Taylor et al., 2015).

For fMRS, the alternative to a blocked design is to use a time-resolved (event-related) design, in which experimental conditions are assigned to individual trials and where the order of trials is often randomised. Meta-analyses of fMRS studies measuring changes in glutamate suggest that compared to blocked designs, time-resolved fMRS designs are associated with larger effect sizes (2-3 times increase in effect size compared to blocked designs) (Mullins, 2018), suggesting that time-resolved designs are indeed more sensitive to changes in neural metabolite concentrations. However, the temporal resolution of time-resolved fMRS remains limited by SNR: using currently available sequences, the SNR of a single MRS spectrum is too low to give a reliable measure of neural metabolite concentrations. For regular MRS or blocked fMRS studies, sufficient SNR is achieved by typically collecting more than one hundred spectra, which are combined into one average spectrum from which neural metabolite concentrations are derived. For time-resolved fMRS, spectra can instead be categorised according to the stimulus type or event, before being combined into one average spectrum per condition to achieve sufficient SNR. This approach has been successfully demonstrated by a handful of studies (Chapter 3.5). Nevertheless, using time-resolved fMRS to measure neural metabolites that reside at relatively low concentration, such as GABA, remains challenging. For this reason specialised 'J-difference edited' sequences that target a specific neural metabolite can be employed. Whether such sequences are optimal for time-resolved fMRS is the topic of the next section.

3.3 Measuring GABA: J-difference versus non-edited sequences

As discussed in Chapter 1 of this thesis, neural inhibition is thought to be an important aspect of neural plasticity and is thought to be implemented by inhibitory interneurons. The activity of inhibitory interneurons cannot be directly measured non-invasively in the human brain. However, MRS can be used to obtain an indirect measure of inhibitory processing, by allowing quantification of GABA, the primary inhibitory neurotransmitter. However, measuring GABA is particularly difficult as its concentration in the brain is low compared to other neural metabolites, and the resonant peaks of GABA overlap with some of these other, more abundant molecules (Andreychenko et al., 2012; Govindaraju et al., 2000; Puts and Edden, 2012). Two main approaches to measuring GABA with MRS are routinely adopted: (1) using a J-difference spectral editing sequence, or (2) using a non-edited sequence. In the following section I will discuss the pros and cons of these two methods, with particular emphasis on their relative suitability for acquiring time-resolved measures of GABA.

The most common approach to measuring GABA is to use the MEscher-Garwood (MEGA) J-difference editing method (Chapter 2.2.2) (Puts and Edden, 2012), where for each measurement of GABA, two spectra are acquired: one regular spectrum, and one with a J-editing pulse which inverts the MRS signal at 3ppm, where a GABA resonance peak is located (Bottomley, 1987; Mescher et al., 1998). Critically, the resonance of GABA at 3ppm is J-coupled to the resonance of GABA at 1.9ppm, and therefore GABA in the edited spectrum resonates differently compared to other molecules. The difference between the regular and edited spectrum thus gives a measure of GABA, which is relatively clean and reliable as most other signal is subtracted out. However, as two measurements are needed for one spectrum, J-difference edited sequences are more prone to subject motion and scanner frequency drift artefacts (Edden et al., 2012). For time-resolved fMRS, the effective reduction in the sampling rate (by a factor of two), is clearly a disadvantage when seeking dynamic, time-resolved measures of neural processes. Finally, the GABA-editing pulse also targets overlapping macromolecules, and the obtained GABA signal is therefore assumed to be contaminated by co-edited macromolecules and referred to as 'GABA+' (Mullins et al., 2014).

GABA can also be measured using non-edited sequences where only one spectrum is needed per measurement. However, this approach is limited by other more abundant molecules which have resonant peaks overlapping with GABA. This is particularly problematic at lower magnetic field strengths. However, with the move to higher field strengths, the spectral peaks of GABA become differentiable from other neural metabolites, including the nearby glutamate peaks. Indeed, GABA has been successfully measured at field strengths of 3 Tesla and up using unedited MRS (Near et al., 2013; Öz et al., 2006; Pradhan et al., 2015). Unedited sequences come with the

benefits of shorter echo times and therefore superior SNR, and simultaneous quantification of the concentrations for ~20 neural metabolites. These properties are advantageous for the application of time-resolved fMRS as it means that more and higher quality data can be gathered in a single acquisition.

In an effort to determine how the differences between sequences affect neural metabolite measurements, several studies have compared edited and non-edited sequences directly with variable results: some found no significant differences between measurements (Hong et al., 2019; Terpstra et al., 2016), while others found that the metabolite of interest (in this case, glutathione) could not be measured without editing (Sanaei Nezhad et al., 2017), or that edited GABA quantification was superior but only after removal of macromolecule signal (C. Chen et al., 2015). In contrast, a recent study comparing reproducibility of glutamate and GABA quantification using edited versus non-edited MRS at 7 Tesla, found that a non-edited sequence with a relatively long echo time provided better reproducibility than either an edited sequence with similar echo time or a non-edited sequence with a shorter echo time (Finkelman et al., 2022). As of yet there are no studies comparing edited and non-edited sequences for time-resolved fMRS. However, given the evidence that there may not be significant differences between edited and unedited sequences (Hong et al., 2019; Terpstra et al., 2016), that reproducibility may be better using unedited sequences (Finkelman et al., 2022) and higher temporal resolution being a strong advantage for time-resolved fMRS studies, unedited MRS sequences may be preferable. This is particularly the case if measurements are performed at higher field strength (i.e. 7 T) with the benefits of higher SNR and higher chemical dispersion (Pradhan et al., 2015).

3.4 Analysis considerations concerning GABA quantification

Having discussed how time-resolved fMRS can be optimised through task-design (3.2) and choice of sequence (3.3), it is also worth noting that certain analysis pipelines may be more optimal than others. This is particularly relevant for protocols designed to detect changes in neural metabolites that are more challenging to quantify, such as GABA. Compared to other more abundant neural metabolites, GABA is difficult to measure due to its relatively low concentration and spectral overlap with other metabolites like glutamate and N-acetylaspartate. To make the quantification of GABA more robust, during the spectral modelling stage GABA is typically assumed to be within a predefined concentration range by spectral fitting algorithms like LCModel. Here, GABA estimates are subject to prior (or 'soft') constraints. In LCModel this range is defined as follows:

$$\frac{[GABA]}{[Big3]} = 0.04 \pm 0.04, \quad [Big3] = [totNAA] + [totCr] + 3[totCho] \tag{3.1}$$

A Gaussian prior is imposed with a mean and standard deviation 0.04 times of the weighted sum of total N-acetylaspartate (totNAA), total creatine (totCr) and total choline (totCho) concentrations (i.e. Big3). These priors keep GABA within a physiologically plausible range. However, they also reduce the variance of the GABA estimate, which may not be optimal for time-resolved fMRS where usually a difference between conditions is evaluated and where variance is therefore of interest.

To test whether including these prior constraints may mask concentration differences in time-resolved GABA between conditions, I used Monte Carlo simulations to generate sets of fMRS data where GABA was artificially added or removed (Figure 3.1A). These simulations show that when prior constraints are removed, GABA estimates are higher and show more variability, while differences in GABA are more reliably detected. Furthermore, changes in GABA can be detected at different SNRs when using the unconstrained approach (Figure 3.1C). By comparison, dynamic sensitivity is reduced when using prior constraints (Figure 3.1B).

These simulations indicate that for the purpose of time-resolved fMRS, foregoing the use of prior constraints may be advantageous. However, absolute metabolite concentrations are higher using this approach as they are not constrained to a physiologically plausible range and thus are harder to interpret. Therefore, rather than assessing absolute concentrations, a difference or ratio between conditions should be the measure of interest.

Time-resolved fMRS studies reported so far have used a variety of mainly 3T sequences, which I will summarise in the next section.

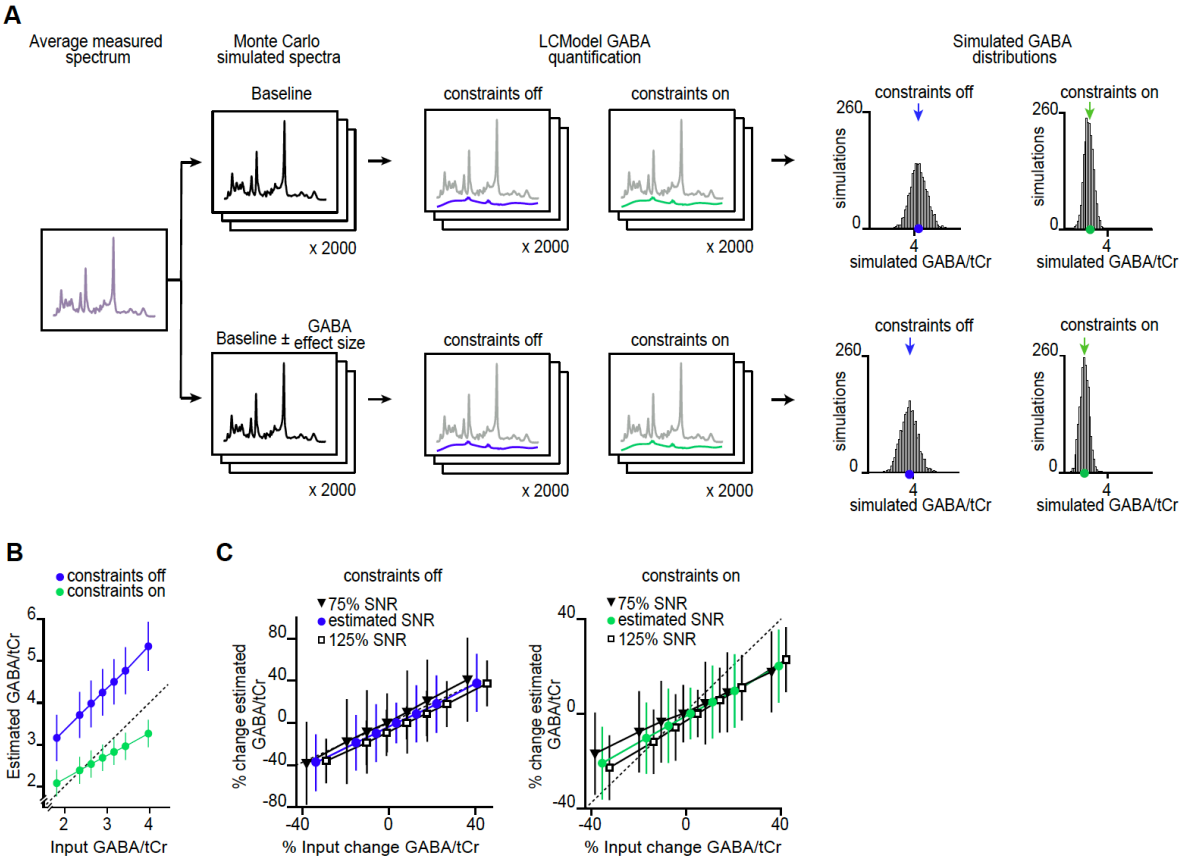


Figure 3.1 Monte Carlo simulations to assess the effect of including/removing prior constraints on GABA. **A** Schematic of the simulation procedure. Sets of spectra were generated with changes in GABA imposed, which included multiples (±0.5, ±1, ±2) of the observed effect size in Chapter 6. Model fitting was performed in LCModel with 'constraints on' (blue) and 'constraints off' (green). **B** For different imposed changes in GABA, the mean ± standard deviation for the resulting GABA:total Creatine (tCR) concentrations (n = 2000) are shown, as estimated with LCModel using 'constraints off' and 'constraints on'. The dashed line represents agreement between the imposed and estimated GABA:tCr concentration. **C** With 'constraints off', the estimated changes closely reflected the imposed changes in GABA at different SNRs. With 'constraint on' the measured changes deviated from the imposed changes at both higher and lower SNRs. These simulations show that dynamic changes in GABA are robustly reflected in the measured GABA concentrations using 'constraints off' at various SNRs. Adapted from (Koolschijn et al., 2021).

3.5 Time-resolved fMRS in the literature

Time-resolved fMRS is still a relatively new method, and for this reason there are currently only a handful of studies reporting results using this approach. However, these studies have investigated a wide range of cognitive processes, demonstrating wide potential applicability.

While currently the majority of studies focus on the role of GABA and glutamate in neural processes, in the first study reporting the use of time-resolved fMRS the role of other neural metabolites, including creatine, in visual processing was investigated (Nishitani, 2003). In this study, the author investigated processing of emotional stimuli in hippocampus using an unedited 3T PRESS sequence. Spectra acquisition was time-locked to presentation of pictures of faces which were presented for 1.8-2.2s, and the author found an increase in creatine/phosphocreatine

and choline-containing compounds upon presentation of unpleasant compared to pleasant faces (Nishitani, 2003). This increase in creatine may reflect increased attention and excitatory activity, as choline is thought to be an indirectly proportional measure of acetylcholine, a neurotransmitter involved in attention and learning (Froemke and Martins, 2011; Hasselmo, 2006). A later study investigating choline by Lindner and colleagues (2017) also used an unedited 3T PRESS sequence and collected spectra in parieto-occipital cortex in response to visuospatial attention shifts. Trials were divided up into an early section (attention condition) and late section (baseline) of 2.5s each, and 'attention' spectra were averaged according to where attention was directed to. Choline concentration was found to be modulated in accordance with attention shifts (Lindner et al., 2017). Thus, these studies show how time-resolved fMRS can be applied to study processes like attention.

The majority of currently published time-resolved fMRS studies have looked at the dynamics of glutamate, the main excitatory neurotransmitter of the nervous system. Gussew and colleagues (2010) studied the glutamate response to pain in insular cortex. Heat stimuli were applied in a cyclical manner over 5 seconds with long inter trial intervals, with 3T PRESS spectra collected at fixed points during the trials. Glutamate concentration in insular cortex was found to increase dramatically (up to 18%) compared to a non-stimulation condition (Gussew et al., 2010). Lally and colleagues (2014) combined time-resolved fMRS with concurrent electro-encephalography (EEG), where participants categorised stimuli as either objects or abstract drawings. Measuring glutamate in lateral occipital complex (LOC) using a 3T PRESS sequence, glutamate levels were found to differ in response to presentation of objects versus abstract visual stimuli, and predicted the evoked gamma-band activity (Lally et al., 2014). Using a stimulus-locked design, Apšvalka et al. (2015) investigated the glutamate response to repeated presentations of stimuli in a repetition suppression paradigm. Regular water-suppressed 3T PRESS acquisitions were interleaved with unsuppressed acquisitions to concurrently measure the BOLD response to visual stimuli in LOC. The fMRS acquisition was locked to the stimulus while the BOLD measurement was acquired in between stimuli. Glutamate showed a relative decrease in concentration upon repeated presentation of stimuli, while novel stimuli induced a relative increase in glutamate (Apšvalka et al., 2015). These studies demonstrate that time-resolved fMRS can be applied to study glutamate dynamics in various neural processes and be combined with other modalities like fMRI and EEG.

Compared to glutamate, studying GABA dynamics using time-resolved fMRS is particularly challenging for the reasons listed earlier in Chapter 3.3. However, there are now two reports of time-resolved GABA measurements. The first study of time-resolved changes in GABA comes from Cleve and colleagues (2015), who measured GABA+ changes in response to heat stimulation in occipital cortex (OC) and anterior cingulate cortex (ACC) using a 3T GABA-edited MEGA-

PRESS sequence. Heat stimuli lasted for 1 s, and pairs of edited and non-edited spectra were acquired during consecutive stimuli, circumventing the issue of lower temporal resolution of an edited sequence. Glx (glutamate + glutamine) was found to increase while GABA+ was found to decrease upon painful stimulation in both OC and ACC (Cleve et al., 2015). In Chapter 6 of this thesis, I apply an interleaved, unedited, fMRI-fMRS 7T sequence in a time-resolved design to study memory recall. Results show an increase in glutamate and decrease in GABA in visual cortex (V1) during successful recall of a visual stimulus. To our knowledge this is the first demonstration of a time-resolved change in the concentration of GABA using an unedited 7T fMRS sequence. These two studies demonstrate the feasibility of time-resolved measures of GABA using fMRS.

While small in number, together these studies demonstrate that time-resolved fMRS can be used to study a wide range of cognitive processes, ranging from pain to attention to associative memory. Furthermore, various study designs are possible, depending on the application: acquisition of fMRS spectra can be time-locked to presentation of a stimulus or spectra can be binned according to the time in the trial they were collected. However, interpreting the temporal dynamics of neural metabolites using fMRS remains non-trivial and is still a matter of discussion in the fMRS community and the focus of the next section.

3.6 Interpreting fMRS-derived changes in metabolite concentrations

Neural metabolites are present in multiple cellular compartments: i.e. cytoplasmic, extracellular and vesicular pools. The spatial resolution of fMRS is insufficient to distinguish between these different neural metabolite pools. Furthermore, the synaptic changes in neural metabolites that accompany neurotransmitter release occur at a timescale that is beyond the temporal resolution of current state-of-the-art fMRS sequences. Given that the spatiotemporal resolution of fMRS does not permit access to fluctuations in neural metabolites at the cellular level and fMRS-derived measures fail to distinguish between different neurotransmitter pools, it remains unclear how task-induced changes in fMRS-derived metabolite concentrations be interpreted. Moreover, interpretation of glutamate and GABA is further complicated by the fact that the release and recycling of glutamate and GABA constitute major metabolic pathways (Bak et al., 2006; Magistretti and Allaman, 2015).

While the relationship between time-resolved fMRS-measures and underlying physiology is not fully understood, MRS is unlikely to be sensitive to changes in the extracellular pools of glutamate and GABA which accompany neurotransmitter release, as the concentration of glutamate and

GABA in the extracellular pools is around 100-fold smaller than in the intracellular pool. In addition, post-mortem studies suggest MRS is not sensitive to intracellular pools that reside in the mitochondria or vesicles (De Graaf and Bovée, 1990; Kauppinen and Williams, 1991). Instead, MRS is considered most sensitive to unconstrained intracellular metabolic pools which reside at sufficiently high concentration in the neuronal cytoplasm (Rae, 2014), at a concentration much higher (~100 fold) than extracellular pools (Myers et al., 2016). Critically, the concentration of glutamate and GABA in these metabolic pools is tightly coupled with neurotransmitter release, which provides the basis for using MRS as a meaningful functional readout of neuronal processing.

A tight coupling between metabolic and neurotransmitter pools can be observed at rest, where a 1:1 relationship is reported between the rate of glutamine-glutamate cycling and neuronal oxidative glucose consumption (Rothman et al., 2003; Shen et al., 1999; Sibson et al., 1998). During sensory stimulation a transient uncoupling has been observed with a short-lived mismatch between glucose utilisation and oxygen consumption (Fox et al., 1988; Fox and Raichle, 1986), particularly during stimulation protocols that alternate between high intensity and quiescent periods (Gjedde et al., 2002). This short-lived uncoupling is only observed during intense and transient sensory stimulation, not during anaesthesia or during certain stimulation protocols. This short-lived uncoupling provides a basis for task-induced changes in MRS measures of glutamate and GABA that reflect functionally relevant activity. In the case of glutamate, the short-lived uncoupling between oxidative metabolism and glutamate-glutamine cycling would lead to an increase in synthesis of glutamate relative to degradation. Opposing effects on GABA as observed in (Cleve et al., 2015) and changes in both glutamate in GABA as reported in Chapter 6 of this thesis, may reflect a transient recalibration or shift in the balance between excitation and inhibition (E/I) at the physiological level.

Nevertheless, the question remains as to whether metabolic changes alone can account for the large task-induced changes (5-12% changes in glutamate (Mullins, 2018) observed using time-resolved fMRS, or whether part of the measured effects reflect changes other than metabolic pool shifts. The possibility has been raised that neural activation drives compartmental shifts between glutamate/GABA states that are more or less visible to MRS. This has been successfully demonstrated using neural spiking models (Lea-Carnall et al., 2021), and is supported by studies that show variation in glutamate visibility to MRS in different compartments (Kauppinen et al., 1994; Kauppinen and Williams, 1991). To further establish the relationship between fMRS-derived neural metabolite measures and compartmental shifts in glutamate and GABA, diffusion-weighted fMRS, a new method that may distinguish neural metabolites in different cellular compartments,

could provide insight (Cao and Wu, 2017). Similarly, preclinical imaging in animal models may provide important mechanistic insight.

More work is needed to establish the functional significance of time-resolved fMRS. Yet, it is worth emphasising that time-resolved fMRS findings to date show remarkable correspondence with dynamic changes in glutamate and GABA predicted by theory, and by empirical studies in both humans and animals (Koolschijn et al., 2021; Vogels and Abbott, 2009).

3.7 Concluding remarks

In this chapter, I have reviewed the use and caveats of time-resolved fMRS as a tool to reveal the neurochemical processes underlying cognition in humans. fMRS in combination with time-resolved task designs has been successfully applied to measure task-induced changes in glutamate and GABA. These dynamic and time-resolved measures of neural metabolites likely reflect transitions to new metabolic steady states, which could indirectly reflect shifts in excitatory versus inhibitory drive at the physiological level. While interpretation of trial-by-trial changes measured using fMRS remains complex, a growing body of evidence supports the notion that fMRS can be used in a time-resolved manner. Furthermore, fMRS is increasingly combined with other modalities like electrophysiology (Lally et al., 2014) and fMRI BOLD (Apšvalka et al., 2015), even in an interleaved manner (Ip et al., 2017). Thus, time-resolved fMRS promises to be a valuable tool to capture real-time, task-relevant changes in neural metabolites, on a timescale relevant for understanding the neural basis of human behaviour.

CHAPTER 4

DISTORTING COGNITIVE MAPS BY ELEVATING NORADRENALINE DURING LEARNING

Memories exist along a trade-off between stability and flexibility. On the one hand, a stable memory system is required for accurate and veridical recall. On the other hand, a flexible system allows for generalisation and making inferences, underlying our ability to support advanced cognition. In this first experimental chapter, I asked whether neuromodulators might be involved in setting the trade-off between stability and flexibility. To this end, I manipulated the concentration of the neuromodulator noradrenaline, which has been associated with decreasing inhibitory tone and increasing neural plasticity in rodents. In humans I predicted that increasing noradrenaline using a pharmacological manipulation would lead to a decrease in cortical GABA, and that this decrease in cortical GABA would be associated with memory (over)generalisation, a sign of distortions in the task-relevant cognitive map. Our data shows that an increase in noradrenaline resulted in a decrease in MRS-derived cortical GABA in the Lateral Occipital Complex (LOC). This effect predicted an increase in neural signatures of (over)generalisation, which in turn predicted a behaviourally equivalent effect. These results suggest that the availability of the neuromodulator noradrenaline during learning influences the trade-off between plasticity and stability of newly acquired cognitive maps.

4.1 Abstract

The relationships between stimuli and events that we experience are thought to be represented in the brain by cognitive maps, which resemble spatial maps of the environment. These cognitive maps assimilate behaviourally relevant distortions, such as overrepresentation of rewarded locations. However, it is currently unclear under what circumstances these distortions arise. Here, we show in humans that increasing the availability of the neuromodulator noradrenaline during learning using the drug atomoxetine is associated with systematic distortions in the behavioural readout of a task-relevant cognitive map. These behavioural distortions are predicted by representational overlap between stimuli observed in the hippocampus, which in turn is predicted by the concentration of neocortical GABA. These data suggest that the availability of noradrenaline during learning may modulate cortical excitability to set the trade-off between a stable yet rigid cognitive map and a more flexible yet distortion-prone cognitive map. These results may help explain why maladaptive overgeneralisation is reported in those clinical populations where levels of noradrenaline are elevated.

4.2 Introduction

Cognitive maps can be defined as the internal organisation of memories formed in response to experiences in the real world. These internal representations provide a basis for memory recall, but also allow agents to take new short-cuts, infer unobserved relationships and generalise information across discrete experiences. Behavioural and physiological studies both provide evidence to support the idea of a cognitive map. At the cellular level, principal cells in the hippocampal formation fire when animals and humans are in a particular location of space. These 'place cells' are considered to provide a neural instantiation of a cognitive map (O'Keefe and Nadel, 1978) which may extend to abstract spaces (Aronov et al., 2017).

Behavioural studies in humans suggest distorted memory maps provide a better explanation for behaviour than objective maps (Bellmund et al., 2020; McNamara, 1986; Warren et al., 2017). For example, measures of distant judgement, direction and map recognition in humans all suggest that cognitive maps reflect a warped version of our lived experience. At the cellular level, spatially tuned place and grid cells in rodents also show evidence for distortions (Butler et al., 2019; Krupic et al., 2015). These distortions to cognitive maps may be adaptive, supporting highly flexible behaviours such as inference and generalisation. Alternatively, these distortions may be maladaptive, leading to unusually intrusive memories or errors in recall. Establishing how distortions are introduced

into memory maps is fundamental for our understanding of both adaptive and maladaptive behaviour. Yet, it remains unclear under what circumstances the cognitive map is more prone to distortions.

One possibility is that distortions in cognitive maps are introduced by fluctuations in neural gain during learning. Brain-wide changes in neural gain are affected by neuromodulators which are elevated by factors that include stimulus salience, arousal and attention. Neuromodulatory tone during learning may therefore set a trade-off between a more stable or more flexible (yet distortion-prone) cognitive map. Consistent with this hypothesis, the location of reward, which is likely accompanied by elevated dopamine, introduces distortions in spatial maps recorded from rodent hippocampus (McNamara et al., 2014). Similarly, distortions in human memory following traumatic events may be explained by elevated levels of neuromodulators, such as noradrenaline, that accompany periods of high stress and arousal. In the extreme, learning in states of stress or high arousal may render a memory maladaptive as observed in Post-Traumatic Stress Disorder where levels of noradrenaline are higher than in healthy controls and predict the severity of clinical symptoms (Kosten et al., 1987; Yehuda et al., 1998). Neuromodulatory tone during learning may therefore affect the propensity for disturbances between learned associations that form a cognitive map.

At a physiological level, a growing literature suggests neuromodulators increase neural gain by regulating the excitability of inhibitory interneurons (Obermayer et al., 2018; Stringer et al., 2016; Tremblay et al., 2016). For example, the neuromodulator acetylcholine reduces stimulus-evoked inhibition (Froemke et al., 2007), while the neuromodulator noradrenaline arguably has a more dramatic effect by decreasing tonic, ongoing inhibition (Martins and Froemke, 2015). These neuromodulatory effects on neural inhibition shift cortical excitatory-inhibitory balance in favour of excitation (Froemke, 2015). This shift towards cortical excitation may facilitate plasticity, strengthening individual memories but also the links between memories. In the extreme, associations may form between memories that have not been experienced together but are either temporally proximal (Cai et al., 2016) or indirectly related (Barron et al., 2020). In this manner, neuromodulators may introduce distortions into cognitive maps. This leads to the hypothesis that by reducing inhibitory tone, elevated noradrenergic tone during learning renders new memories prone to distortion.

To test this hypothesis, in humans we asked whether learning during periods of elevated noradrenaline introduces distortions into cognitive maps. We designed a task where participants were required to learn pairwise associations between visual stimuli, which were organised

according to a ring structure which constituted a non-spatial map. Participants were given a single oral dose of the noradrenaline reuptake inhibitor atomoxetine to transiently elevate noradrenaline in a drug group ('ATX'), with comparison to a placebo control group ('PLC'). After learning we then used functional magnetic resonance imaging (fMRI) to index the representational overlap between different stimuli on the ring and we monitored the effect of noradrenaline on neural excitability using Magnetic Resonance Spectroscopy (MRS).

We show that elevating noradrenaline during learning introduces systematic distortions in a behavioural readout of participants' cognitive map. This behavioural result is predicted by the equivalent representational overlap between stimuli in the hippocampus and in turn may be explained by a reduction in neocortical GABA that accompanies a single dose of atomoxetine. Together these findings suggest that learning under elevated noradrenaline increases the propensity for spurious mnemonic associations and introduces systematic distortions in cognitive maps. Such distortions may allow the brain to readily build inferences between salient memories to support flexible behaviour. Conversely, elevated noradrenaline may underlie maladaptive overgeneralisation, as reported in clinical populations.

4.3 Methods

4.3.1 Participants and double blinding

44 healthy volunteers participated in this randomised double-blind, placebo-controlled, between-subject study. Participants were randomly assigned to 2 sex-matched groups: the treatment group who received a single dose of 10 mg atomoxetine (ATX group mean age: 23.9 ± 5.23 yrs) or the placebo group (PLC group mean age: 25.2 ± 4.80 yrs) (mean ± SEM). All participants were given an over-encapsulated capsule, which contained either atomoxetine or an empty capsule. Randomisation was performed by a researcher not involved in recruitment and data collection. The drugs were administered by the experimenter (RK), who was blind to the drug condition. All experiments were approved by the University of Oxford ethics committee (reference number R60579/RE004). All participants gave informed written consent.

For one participant in the ATX group and 2 participants in the PLC group, no MRS data was collected in V1. Due to low data quality 7 participants (ATX: 3 PLC: 4) were excluded from MRS LOC data analysis, and 3 participants (ATX: 1 PLC: 2) were excluded from MRS V1 data analysis. For one participant in the placebo group only 3 out of 4 fMRI blocks were collected due to a fault of the scanner. For 10 participants (ATX: 7 PLC: 3), pupillometry data could not be analysed as

the data from the eye tracker was too noisy. One participant in the placebo was excluded from the fMRI analyses due to excessive movement in the scanner. For another participant in the placebo group, only one scan block was completed due to a fault of the MRI scanner.

4.3.2 Behavioural training

On the first day of the experiment, participants (n=44) received either a single dose of atomoxetine or placebo, followed by a 90 minute waiting period (Figure 4.1D). The waiting period was chosen to achieve a neuromodulatory effect during the learning task and subsequent MRI scan, based on the drug's pharmacodynamics (Sauer et al., 2005). Following the waiting period, participants learned sets of context dependent associations between bird stimuli (Figure 4.1A,B). The contextual cue for each association comprised a different room which included a sofa (Figure 4.1A,B). Importantly, the bird stimuli were always presented in the same location in their context and never on or near the contextual stimulus, to ensure the participants would be unlikely to explicitly link the birds with the sofas. To explicitly learn the associations, participants perform a 3-alternative forced choice task with feedback (Figure 4.1E). Together, the context dependent bird associations formed a symmetrical ring structure (Figure 4.1C). This ring structure was optimised for measuring inhibitory engrams using fMRI cross stimulus suppression, where each context dependent association could be used as a control for another context dependent association on the other side of the ring. After participants scored >90% correct on the learning task for two consecutive blocks (22 trials each), participants completed the inference test (Figure 4.1F). In this 3-alternative forced choice test (44 trials), participants were asked which stimulus shares an association with the probe bird (e.g. bird '1' and bird '3') in the absence of contextual cues or feedback, giving participants the opportunity to make trajectories around the ring structure.

Four days later, on Day 5, participants were tested on the explicitly and implicitly learned associations. In the explicit memory test (Figure 4.2A), participants completed the same 3-alternative forced choice task as during learning, where they identified the stimulus that was directly associated to the probe, in the absence of contextual cues or feedback. Subsequently, participants again completed the inference test they had encountered on the first day (Figure 4.1F). Furthermore, participants completed the implicit memory test (Figure 4.2B, 264 trials divided over 6 blocks). During each trial of the implicit memory test, one bird stimulus (probe) and six contextual cues (sofas) were presented in a six-alternative forced choice paradigm without feedback. Here, participants had the opportunity to make overgeneralisation errors, where they erroneously linked a stimulus to a neighbouring contextual cue (Figure 4.2C, e.g. bird '1' and sofa

'B'). Overgeneralisation errors were defined as the percentage of trials where a foil error was made, subtracted by the percentage of trials where a non-foil error was made.

4.3.3 MRS data collection and analysis

On the first day of the study, MRS spectra were acquired using a MEscher-GArwood Point RESolved Spectroscopy (MEGA-PRESS) sequence in two 2 x 2 x 2 cm3 VOIs, with TE = 68 ms, TR = 1.5 s, flip angle 90 deg, ON editing pulse 1.9 ppm, OFF editing pulse 7.5 ppm. Unsuppressed water signal was collected for internal water referencing. The first volume of interest (VOI) was positioned centered bilaterally on the calcarine sulcus in visual cortex (V1). The second VOI was positioned in LOC. For each measurement 320 spectra were acquired (160 edit-ON and 160 edit-OFF), resulting in an acquisition time of about 8 min for each VOI. B_0 shimming was performed using a GRE-SHIM (FOV =216 mm2; TR= 482 ms; TE1/2 = 4.92/7.38 ms; slice thickness 2 mm; flip angle 46 deg; slices 49; scan time 105 s).

Glx and GABA+ concentrations were quantified using Gannet version 3.1.4 (Edden et al., 2014) applied to the raw Siemens 'twix' data. Water was used as an internal concentration reference. Gannet's standard preprocessing pipeline was used, which includes frequency and phase correction by spectral registration and line broadening. Edited spectra were generated by averaging and subtraction of individual edit-ON and edit-OFF spectra. The editing approach does not only target GABA but also other macromolecules at 3ppm, and therefore GABA+ (GABA + macromolecules) concentrations are reported. Furthermore, the editing approach also targets nearby glutamate/glutamine resonances, allowing for Glx concentration estimates to be made. Metabolite concentrations were relaxation and tissue-corrected (Gasparovic et al., 2006). Grey matter, white matter and CSF tissue fractions for determining tissue-corrected concentrations were obtained for both voxels using SPM12.

Reliable model fits were achieved for 38 out of 41 V1 acquisitions and for 36 out of 43 LOC acquisitions. In the remaining acquisitions GABA+ was unreliable or inestimable due to lipid contamination or low SNR. Data quality was quantified using the signal-to-noise ratio (SNR) and full-width-at-half-max (FWHM) of N-acetylaspartate (NAA) and fit error of the GABA+ peak provided by Gannet (Supplementary Figure 4.3).

Group average GABA+ and Glx concentrations were compared using two-tailed permutation tests. A null distribution of the expected difference was generated by permuting group labels (ATX, PLC) 10,000 times using MATLAB's random number generator.

4.3.4 fMRI scan task

During the fMRI scan, participant viewed the 11 bird stimuli during 2 scan blocks, and the 11 bird stimuli interleaved with the 11 sofa contextual stimuli during 2 scan blocks. Stimuli were presented via a computer monitor inside the scanner bore. On each trial, stimuli were presented for 1000 ms (Figure 4.3A). The inter-trial interval was selected from a truncated gamma distribution with a mean of 2.3 s, minimum of 1.3 s and maximum of 14 s. To control for potential confounding effects of expectation suppression (Summerfield et al., 2008), all possible transitions between pairs of non-repeating stimuli were presented once in each block, in a fully randomised order. Participants performed a task incidental to the contrast of interest which involved identifying whether the presented stimuli were familiar or "oddball". Oddball stimuli were transformed or differently coloured versions of the stimuli encountered during training, and were randomly inserted into 9% of trials. Participants were instructed to press a button on an MR compatible button box using their right index finger when they identified "oddball" stimuli, but not if the stimulus was familiar. Participants received a monetary reward for each correct oddball identification. No feedback was given other than the total monetary reward after each block. Blocks with only bird stimuli lasted 10 minutes and included 133 trials. Blocks where bird stimuli were interleaved with contextual cues lasted 19 minutes and included 265 trials. All stimuli were presented 12 times.

4.3.5 Pupillometry acquisition, preprocessing, and analysis

To assess whether elevated noradrenaline was associated with altered pupil responses in response to surprising stimuli, we collected pupillometry data during the fMRI scan task. We measured pupil size using an EyeLink eyetracker (SR Research) sampled at 500 Hz. The pupillometry data was preprocessed using standard methods. We identified and removed blinks by interpolating the values just before and after the identified blinks. Trials on which the pupil measurement was lost for a period of >1000ms were removed from the analysis. Participants were not excluded if more than 50% trials had to be excluded. This resulted in the inclusion of 34 out of 44 participants, with an average of 672.50 ± 17.69 'regular' trials and 58.82 ± 1.36 'oddball' trials (mean ± SEM).

Pupil sizes were normalised across each scan block by expressing them in terms of the percentage difference from the overall main pupil size across the block. Baseline pupil size for each trial was defined as the mean of the normalised pupil size in the 250 ms window prior to

stimulus presentation. To assess the effect of surprise, for each participant we computed the difference between average pupil response for 'oddball' vs 'regular' stimuli. To assess the difference between the ATX and PLC group, we used cluster-based significance approach where we computed the size threshold for a significant cluster by permuting subject labels x10,000 for each timepoint, and for each iteration noting the maximum size of a difference cluster (Hunt et al., 2013; Hauser et al., 2015).

4.3.6 fMRI imaging protocol

The fMRI scan task was performed inside a 3 T Magnetom Prisma (Siemens) using a 32-channel head an neck coil (Nova Medical Inc, Wilmington, MA) at the Wellcome Centre for Integrative Neuroimaging (University of Oxford). fMRI data was acquired using a multiband echo-planar imaging (EPI) sequence, with multiband factor 3, TR = 1.235 s, TE = 20 ms, flip angle = 65 degrees, field of view = 216 mm, and a voxel resolution of $2\times2\times2$ mm^3. All scans were of axial orientation angled to hippocampus, covering the whole brain. For each participant, a T1-weighted structural image was acquired to correct for geometric distortion and perform co-registration between EPIs, consisting of 192 1 mm axial slices, in-plane resolution of 1×1 mm^2, TR= 1.9 s, TE = 3.97 ms, and field of view = 192 mm. A field map with dual echo-time images was also acquired (TE1 = 4.92 ms, TE2 = 7.38 ms, whole-brain coverage, voxel size $2\times2\times2$ mm^3). Physiological measures (cardiac pulse and thoracic movement) were collected during the scan. Cardiac pulse was recorded using an MRI-compatible pulse oximeter and thoracic movement was monitored using a custom-made pneumatic belt positioned around the abdomen.

4.3.7 fMRI preprocessing and GLMs

Preprocessing of fMRI data was carried out in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). For the 43 participants included in the fMRI analysis, images were corrected for signal bias to account for magnetic field inhomogeneities, realigned to the first volume of the block, corrected for distortion using fieldmaps, normalised to a standard EPI template and smoothed using a 5 mm full-width at half maximum Gaussian kernel. To remove low frequency noise from the preprocessed data, a high-pass filter was applied to the data using SPM12's default settings. For each participant and each scan block, the resulting fMRI data was analysed in an event-related manner using two General Linear Models (GLMs). One GLM was applied to the data from the two scan blocks where only bird stimuli were presented, and one GLM was applied to the data from the two scan blocks were bird stimuli were interleaved with contextual cues. In addition to

the explanatory variables (EVs) of interest (described below), six additional scan-to-scan motion parameters produced during realignment were included in the GLM as nuisance regressors to account for motion-related artefacts in each task block.

In the first GLM, used for scan blocks where only bird stimuli were presented, 7 EVs were included per scan block. The first 5 EVs accounted for the possible link distance between the presented stimulus and the preceding stimulus, measured along the ring structure (1-link to 5-link, Figure 4.1C). The remaining 2 EVs accounted for detected and undetected 'oddball' trials, respectively. All EVs were then convolved with the hemodynamic response function.

In the second GLM, used for scan blocks where bird stimuli were interleaved with contextual cues, 16 EVs were included per scan block. The first 6 EVs accounted for the possible link distance between the presented bird stimulus and the preceding contextual stimulus, measured along the ring structure (1-link to 6-link, Figure 4.1C). Similarly, the next 6 EVs accounted for the possible link distance between the presented contextual cue and the preceding bird stimulus (1-link to 6-link). The final 4 EVs accounted for detected and undetected bird 'oddball' trials and detected and undetected contextual cue 'oddball' trials.

4.3.8 Univariate fMRI analysis and statistics

Using the output of the first GLM, we measured cross-stimulus suppression as an index for expression of directly associated bird stimuli (Figure 4.5A). The contrast of interest involved comparing fMRI BOLD signal for trials where the preceding stimulus was distal on the ring (Figure 4.1C, at least 4 links away, e.g. bird '1' preceded by bird '5') with fMRI BOLD for trials with pairs of stimuli that were directly linked on the ring structure (e.g. bird '1' preceded by bird '2').

Next, we used the second GLM to measure cross-stimulus suppression as an index for a 'spread of association' effect (Figure 4.2C). The contrast of interest involved comparing fMRI BOLD signal for trials where the preceding contextual cue was distal on the ring (at least 4 links away, e.g. bird '1' preceded by sofa 'E') with fMRI BOLD for trials where the preceding contextual cue was indirectly linked to the presented stimulus (e.g. bird '1' preceded by bird 'B'). The resulting contrast images for all participants were entered into a second-level random effects 'group' analysis. We set the cluster-defining threshold to $p < 0.001$ uncorrected before using whole-brain family wise error (FWE) to correct for multiple comparisons, with the significance level defined as $p < 0.05$ (Figure 4.7).

4.3.9 Assessing the relationship between fMRI, behaviour, and GABA+

To assess the relationship between XSS effects and overgeneralisation errors across the entire imaged brain volume, we repeated the second-level random effects 'group' analysis using the output of the first GLM, but now included group-level covariates for overgeneralisation errors (i.e. Figure 4.2F). To identify brain regions where the memory XSS index predicted overgeneralisation errors, for each group we contrasted the EV of the behavioural covariate to generate a single contrast to test statistical significance. To test the effect of group, we entered the EVs for group (2) and EVs for covariance (2, one for ATX and one for PLC), and contrasted the covariance EVs to generate a group difference contrast. We set the cluster-defining threshold to $p < 0.001$ uncorrected before using whole-brain FWE to correct for multiple comparisons, with the significance level defined as $p < 0.05$ (Figure 4.6).

Next, to assess the relationship between hippocampal BOLD XSS effects and the concentration of GABA+ in LOC we used an anatomical ROI for hippocampus for the analysis in Figure 4.5E and a hippocampal ROI based on the contrast Figure 4.5B for the analysis in Figure 4.7B. Capitalising on the variance across participants, the relationship between XSS measures for memory and 'spread of association' within these ROIs was compared with GABA+ using a Spearman rank correlation (Figure 4.5E, Figure 4.7B).

4.4 Results

4.4.1 Experimental design

To test our hypothesis that elevated noradrenaline during learning introduces systematic distortions in a cognitive map, we designed a task that required participants to learn a total of 11 pairwise associations. Each pairwise association involved a pairing between two bird stimuli (Figure 4.1A,B). We designed the task such that the underlying relational structure between the pairwise associations was susceptible to distortion. Thus, each bird stimulus was paired with two other bird stimuli, and the set of associations could be arranged into a ring structure (Figure 4.1C), though participants were not informed of this. Furthermore, during learning each pair of bird stimuli was presented in a specific context indicated by a living room interior (Figure 4.1A-B).

Participants learned the pairwise associations between birds using a three-alternative forced choice task ('learning task', Figure 4.1D,E). Knowledge of the specific contextual cue that accompanied each pair of bird stimuli remained implicit as participants were not tested on these

relationships during learning. To promote formation of memories within a cohesive map-like structure (Carpenter and Schacter, 2017), after learning participants were given the opportunity to infer indirect relationships between bird stimuli that share a common association, thereby making trajectories around the ring structure ('inference task', without feedback; Figure 4.1F). The ring topology further allowed us to test whether participants represent the task using a distorted cognitive map that includes a 'spread of association' around the ring. We predicted that learning the associations during a period of elevated noradrenaline should make participants susceptible to this spread of association.

To manipulate noradrenaline during learning we used a between-subject double-blind randomised placebo-controlled design. At the start of the experiment, participants were given either a single dose of the noradrenergic reuptake inhibitor, atomoxetine, or a placebo. Atomoxetine is a drug previously shown to increase levels of noradrenaline in both neocortical brain areas and the hippocampus (Bymaster et al., 2002; Tzavara et al., 2006). Participants could not tell whether they received atomoxetine or placebo (Supplementary Figure 4.1). After receiving the atomoxetine/placebo, participants waited 90 minutes before commencing the task.

All participants learned the pairwise associations with high accuracy and there was no significant difference in learning between the atomoxetine ('ATX') and placebo ('PLC') groups (Figure 4.1G,H). Moreover, there was no significant difference in performance between ATX and PLC on the inference task (Figure 4.1I).

To investigate whether learning under elevated noradrenaline increases the incidence of memory distortions in the cognitive map, participants underwent an MRI scan after learning was complete (Figure 4.1D). Using fMRI we indexed the similarity between subvoxel neuronal activity patterns for different stimuli. We combined this approach with pupillometry and MRS, the latter being a technique that measures local cortical gamma-aminobutyric acid (GABA) and glutamate concentration. In addition, four days after the scan session (day 5, Figure 4.1D) participants returned to perform behavioural tasks that allowed us to measure the spread of association between memories, when the effects of atomoxetine had washed out (Sauer et al., 2005).

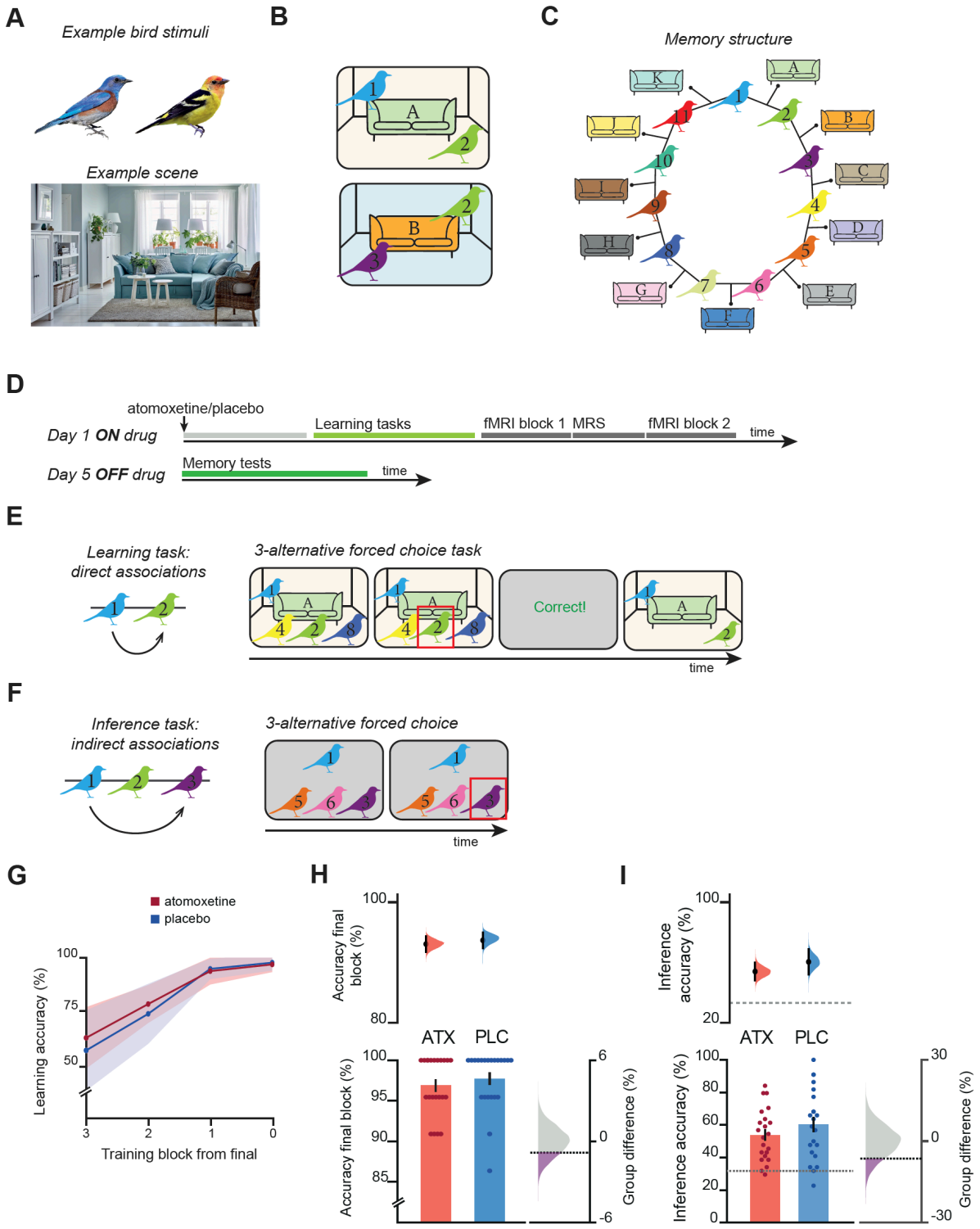


Figure 4.1 Experimental design

A Example stimuli and context used in the experiment. Stimuli were images of birds and contexts were living room scenes containing a sofa. 11 bird and living rooms were used. **B** Schematic of the context dependent associations. Each stimulus was associated with two other stimuli, and each association was linked to a specific context. **C** Participants learned 11 pairwise associations which could be arranged as a ring structure, though this information was not shared with the participants. The spikes represent the contextual cues. **D** Schematic of the experimental protocol. On Day 1, participants received either atomoxetine or placebo. After a 90 min waiting period participants completed the learning task, inference test, and underwent an MRI scan. On day 5, participants completed memory tests for the explicit and implicit associations encountered on Day 1. **E** To learn the stimulus associations, on day 1 participants completed a three-alternative forced choice learning task which included contexts and feedback. **F** After completing the learning task, participants performed an associative inference test, where they were required

to infer which bird stimulus was indirectly associated to the probe in a three-alternative forced choice task in the absence of feedback **G** Participants performed the learning task until they successfully remembered 90% of the associations in two runs. **H,I** Upper: Bootstrap-coupled estimation (DABEST) plots. Black dot, mean; black ticks, 95% confidence interval; filled curve, sampling error distribution. Lower left: memory accuracy (mean +/- SEM). Lower right: null distribution of the group differences generated by permuting subject labels, dotted line: true group difference. **H** All participants successfully completed the learning task. There was no significant difference between groups (ATX – PLC: permutation test: p=0.206). **I** There was no significant difference between inference task performance between the groups (ATX – PLC: permutation test: p=0.128).

4.4.2 Elevated noradrenaline during learning increases behavioural measures of memory distortions

To assess behavioural evidence for a spread of association between memories on day 5, we tested memory for both the learned pairwise associations ('explicit memory test', Figure 4.2A) and the incidental pairings between each bird stimulus and the contextual cues ('implicit memory test', Figure 4.2B). To test memory for incidental pairings we assessed the relationship between the bird stimuli and images of sofas taken from the background interiors that formed the contextual cues. Importantly, throughout the task the bird stimuli were always presented in the same location within their background interior and never on or proximal to the sofas. Moreover, the implicit memory test provided an opportunity to assess whether participants erroneously made a link between the probed bird stimulus (e.g. bird '1') and the neighbouring contextual cue on the ring, namely the other contextual cue associated with the linking bird stimulus (e.g. bird '2'). We predicted that increased noradrenaline during learning would increase the propensity of subjects to form erroneous associative links between stimuli and contextual cues that are indirectly associated (e.g. bird '1' and sofa 'B' or 'J' rather than sofa 'A'), thus showing a spread of association in the underlying memory map (Figure 4.2C).

Behavioural performance on day 5 indicated no significant group difference in explicit memory accuracy for the pairwise associations that were learned on day 1 (ATX vs. PLC, 'explicit memory accuracy', Figure 4.2D). Similarly, there was no significant group difference in inference accuracy tested on day 5 (ATX vs. PLC, 'inference accuracy', Supplementary Figure 4.2) and no significant difference in overall implicit memory accuracy for relationships between the birds and their contextual cues (ATX vs. PLC, 'implicit memory accuracy', Figure 4.2E).

To obtain a read-out of participants' tendency to form erroneous links between stimuli and indirectly related contextual cues, we analysed the error trials in the implicit memory test, where participants made errors on 75.4% of trials on average. Consistent with our prediction, during the implicit memory test participants in the ATX group, but not PLC, were significantly more likely to recall erroneous associative links between stimuli and contextual cues that are only indirectly associated than would be expected by chance (Figure 4.2F). Furthermore, in the ATX group, but

not in the PLC group, this behavioural evidence for a spread of association in the underlying memory map was predicted by performance on the inference test performed on day 1 (Figure 4.2G). These results suggest that learning under elevated noradrenaline increases the spread of associations between stimuli that were never presented together but are located nearby in the cognitive map. Furthermore, making inferences between stimuli during periods of elevated noradrenaline predicts this subsequent spread of association between memories.

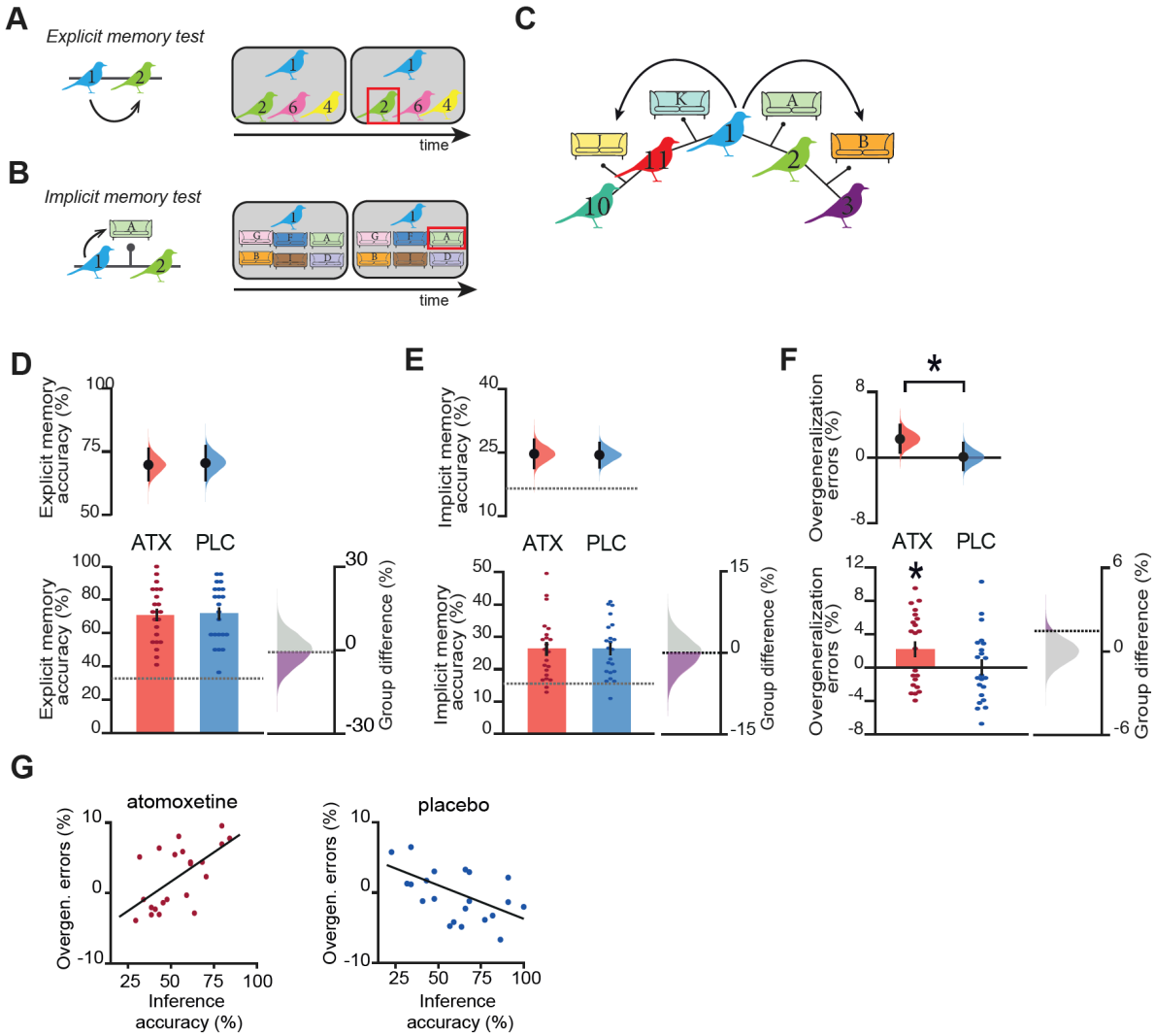


Figure 4.2 Noradrenergic stimulation during learning increases overgeneralisation
A Example trial of the explicit memory test performed by participants on day 5, where the participants were tested on the associations learned on day 1. **B** Example trial of the implicit memory test performed by participants on day 5, where the participants were required to recall which contextual cue (sofa) was presented with the probe during learning. **C** Illustration of the 'spread of association' effect, i.e. the formation of links between stimuli and contextual cues that are only indirectly associated. **D-F** Upper: Bootstrap-coupled estimation (DABEST) plots. Black dot, mean; black ticks, 95% confidence interval; filled curve, sampling error distribution. Lower left: memory accuracy (mean +/- SEM). Lower right: null distribution of the group differences generated by permuting subject labels, dotted line: true group difference. **D** In the explicit memory test, there was no significant difference between groups (ATX – PLC: permutation test: p=0.421). **E** In the implicit memory test, there was no significant difference in overall accuracy between groups (ATX – PLC: permutation test: p=0.498). **F** During the implicit memory test, participants made errors on 75.4% of the trials (ATX: 75.3%, PLC: 75.6%), with the opportunity to make 'overgeneralisation errors' defined as those errors where the contextual stimulus indirectly linked to the probe was chosen (C). The ATX group made significant amounts of overgeneralisation errors (p=0.014, effect size computed

from 10,000 bias-corrected bootstrapped resamples (Ho et al., 2019)) and significantly more overgeneralisation errors compared to PLC (ATX – PLC: permutation test p=0.048). **G** Left: Performance on the inference test on day 1 positively predicted the percentage of overgeneralisation errors in the ATX group (Spearman correlation: r_{20} = 0.61 p=0.002). Right: In the PLC group this relationship was reverse, i.e. performance on the inference test negatively predicted the number of overgeneralisation errors (Spearman correlation: r_{20}=-0.49, p=0.02).

4.4.3 Elevated noradrenaline modulates pupil response to surprising stimuli

To identify the neuronal mechanism underlying this spread of association, we first verified that atomoxetine alters established physiological markers of noradrenaline. Work in animals and humans suggest pupil diameter provides an index for locus coeruleus activity (Gilzenrat et al., 2010; Rajkowski et al., 1993; Samuels and Szabadi, 2008) and changes in pupil size have been shown to reflect surprise (Nassar et al., 2012; O'Reilly et al., 2013; Preuschoff et al., 2011). Here we took advantage of this physiological marker for noradrenaline by collecting pupillometry data. Pupillometry data were acquired after the learning task during the MRI scan, while participants performed an "oddball" detection task. During the oddball detection task participants were presented with familiar stimuli observed during learning and were required to detect oddball stimuli that made up ~9% of trials (Figure 4.3A). We compared participants' pupil response to oddball stimuli versus familiar stimuli. We hypothesised that the increase in neural gain due to increased availability of noradrenaline in the ATX group would amplify the increase in pupil size to surprising versus familiar stimuli. In both the ATX and PLC we observed an increase in pupil dilation response to oddball stimuli. Importantly, and consistent with previous findings that suggest pupil dilation reflects noradrenaline levels (Rajkowski et al., 1993; Samuels and Szabadi, 2008), this pupil response to surprise was greater in the ATX group ~6-10 s after stimulus onset (Figure 4.3B). This result confirms that the single dose of atomoxetine induces changes in physiological markers of noradrenergic arousal.

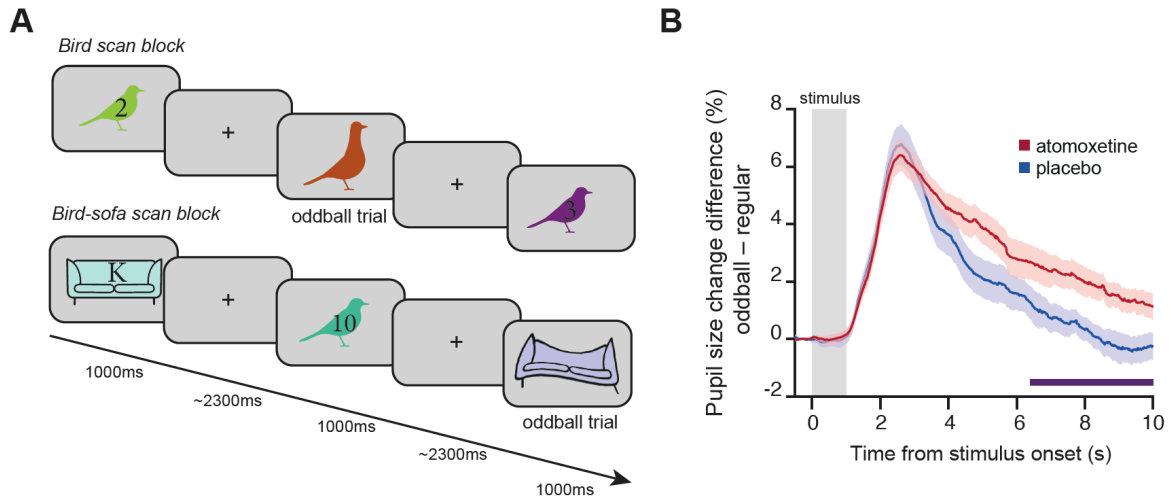


Figure 4.3 Elevated noradrenaline increases pupil surprise response
A During the scan task, participants observed bird and sofa stimuli encountered during the learning task presented against a grey background. Participants were instructed to detect "oddball" stimuli, encountered in ~9% of trials.

Oddball stimuli were warped or differently coloured versions of the familiar stimuli. All possible transitions between stimuli were presented once in a randomised order. Participants completed 4 fMRI scan blocks (Figure 1D). Two of the blocks included bird stimuli only, and in the other two blocks bird and contextual cues were interleaved. **B** Oddball stimuli induced an increase in pupil response compared to familiar stimuli. Between 6 and 10 s after stimulus onset, there was a significant group difference in the magnitude of the increased pupil size response effect (purple horizontal line: cluster-corrected significant ATX – PLC; permutation test corrected for multiple comparisons: p=0.025).

4.4.4 Elevated noradrenaline decreases inhibitory tone in neocortex

Having confirmed that our noradrenergic manipulation induces expected changes in pupil response, we next asked whether a single dose of atomoxetine also changes non-invasive measures of neural excitability. The majority of noradrenergic projections to cortex originate in the locus coeruleus (LC) and stimulation of LC in rodents induces a decrease in cortical inhibitory tone (Martins and Froemke, 2015). In humans, we therefore predicted that elevating levels of noradrenaline using a single dose of atomoxetine would induce a reduction in cortical inhibition. To measure cortical inhibition non-invasively, we quantified the concentration of cortical GABA+ (GABA and macromolecules) using MRS, a non-invasive method used *in vivo* to measure the relative concentration of target metabolites in the brain. We measured Glx (glutamate and glutamine) and GABA in two 2 x 2 x 2 cm^3 voxels, positioned in lateral occipital complex (LOC) and primary visual cortex (V1), cortical regions involved in visual object recognition (Grill-Spector et al., 2001, 1999).

In line with previous evidence from rodents (Martins and Froemke, 2015), GABA+ measured from a voxel positioned in the LOC (Figure 4.4A) was significantly reduced in the ATX group compared to the PLC group (Figure 4.4B). There was no difference in the concentration of Glx (Figure 4C). When using measures of both Glx and GABA+ to estimate E/I ratio, a significant increase in E/I ratio was observed in ATX compared to PLC (Glx:GABA+, Figure 4.4D). Standard quality metrics indicated that there were no differences in data quality between the ATX and PLC groups (Supplementary Figure 4.3). These results suggest that elevating noradrenaline using a single dose of atomoxetine reduces cortical inhibition, thereby increasing cortical excitability. However, no significant difference in GABA+ or E/I ratio was observed in the voxel positioned in V1 (Supplementary Figure 4.4).

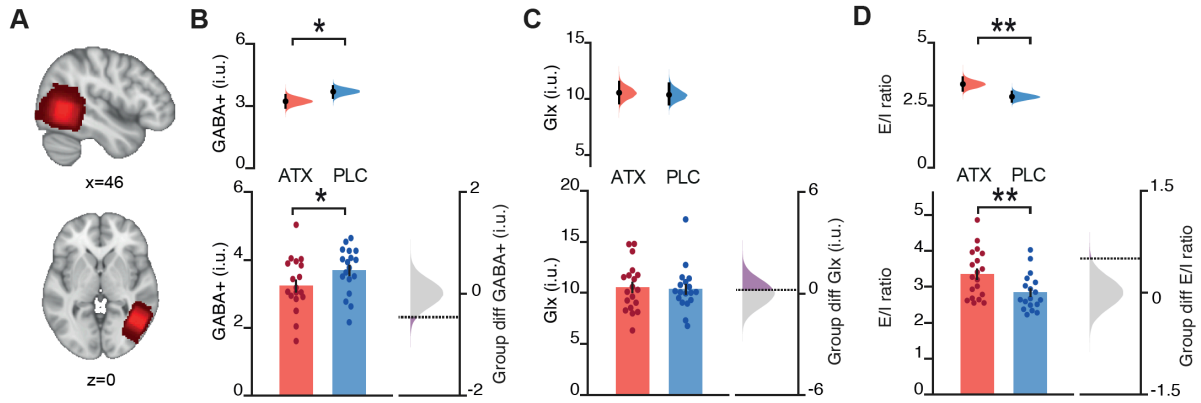


Figure 4.4 Elevated noradrenaline locally decreases inhibitory tone in neocortex
A Anatomical location of the 2 x 2 x 2 cm^3 MRS voxel, positioned in LOC. Cumulative map across 37 participants. **B-D** We used MRS to quantify the effect of atomoxetine on excitatory and inhibitory tone in neocortex, ~3.5 h after drug intake. Upper: Bootstrap-coupled estimation (DABEST) plots. Black dot, mean; black ticks, 95% confidence interval; filled curve, sampling error distribution. Lower left: metabolite concentrations and E/I ratio (mean +/- SEM). Lower right: null distribution of the group differences generated by permuting subject labels, dotted line: true group difference. **B** A significant group difference was observed in the concentration of GABA+ (GABA+ ATX – PLC: permutation test p=0.034). **C** No significant group difference was observed in the concentration of Glx (Glx ATX – PLC: permutation test p=0.402). **D** There was a significant group difference in E/I ratio, defined as the ratio of Glx:GABA+ (E/I ratio ATX – PLC: permutation test p=0.007).

4.4.5 Hippocampus represents learned stimulus associations

We hypothesised that as a result of new learning, neuronal activity patterns for stimuli on the ring should reflect the learned associations. To test this prediction, we used fMRI repetition suppression. fMRI repetition suppression relies on the experimental observation that neurons show a relative suppression in their activity when presented with repeated stimuli to which they are responsive (Miller et al., 1991; Sawamura et al., 2006). This feature cannot only be used to access sub-voxel representations for single stimuli (Grill-Spector et al., 2006), but also to measure relative coactivation or overlap between representations coding for different stimuli (Barron et al., 2016a), termed cross-stimulus suppression (XSS).

Here, to measure the relative coactivation between associated bird stimuli using XSS, during the fMRI scan stimuli were presented in a pseudo-randomised order. Participants were required to perform an oddball detection task, as described above (Figure 4.3A). The presentation of each stimulus was categorised using the stimulus that came immediately before (Figure 4.5A). The ring topology of the learned associations provided an efficient way to ensure that each stimulus contributed to both trials where suppression was expected (stimulus pairs that were proximal on the ring) and control trials where suppression was not expected (stimulus pairs that were distal on the ring).

To assess fMRI XSS for associated bird stimuli, we compared the BOLD response to bird stimuli that were preceded by a directly associated bird stimulus (e.g. bird '2' followed by bird '1')

with control trials where no suppression effects were expected (e.g. bird '5' followed by bird '1') (Figure 4.5A). In both the ATX and PLC groups we observed significant fMRI XSS between associated stimuli in the hippocampus (Figure 4.5B,C), indicating that hippocampal representations for associated bird stimuli are co-active. Per group analyses revealed that this effect in hippocampus was significant in the ATX group only (Figure 4.5D), but not PLC (Supplementary Figure 4.5A), with no significant difference between the two groups (Supplementary Figure 4.5B). Overall, this suggests learning under elevated noradrenaline leads to significant coactivation between representations of paired stimuli in hippocampus.

Next, we asked whether inhibitory tone might underlie this increase in coactivation between paired associates following learning. If the representational overlap between paired stimuli reflects an increase in neuronal excitability following learning, increased coactivation between paired stimuli as indexed using fMRI XSS should be associated with lower inhibitory tone. In line with this prediction, taking advantage of variance observed across all participants, we observed a negative relationship between GABA+ measured in LOC and the coactivation between paired associates in hippocampus (Figure 4.5E). Together, these results show that representations of learned stimulus associations are co-active in hippocampus, and this effect is observed when noradrenaline is elevated. Furthermore, neocortical inhibitory tone predicts coactivation between paired associates in hippocampus, suggesting that this coactivation is related to neural excitability of cortical inputs.

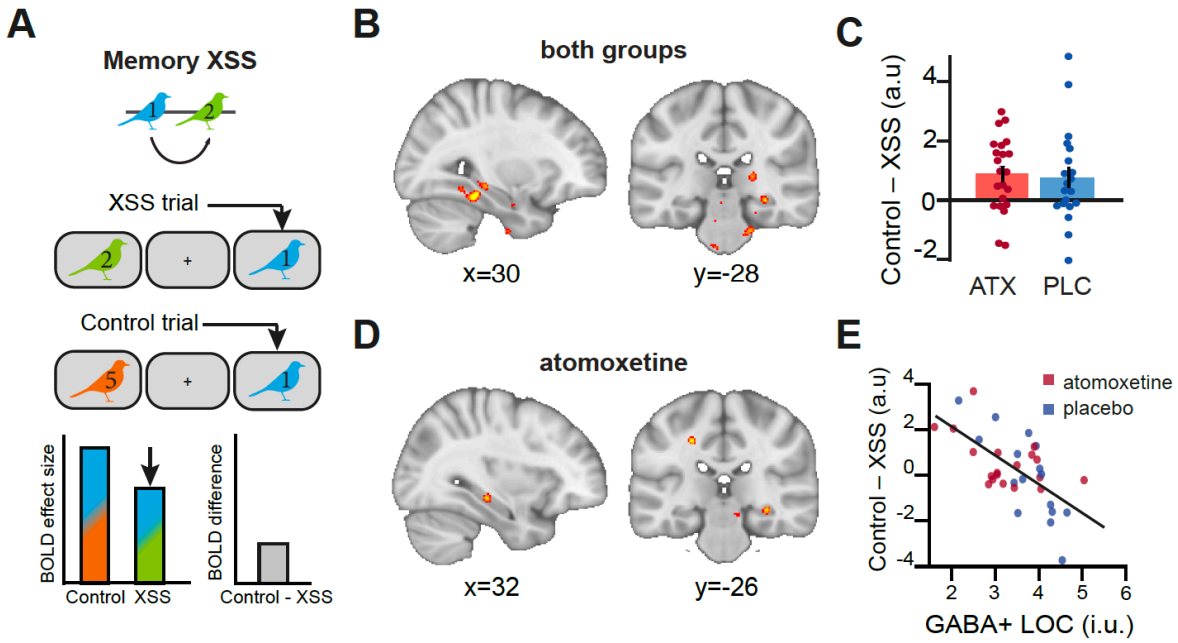


Figure 4.5 Hippocampus represents stimulus relationships and hippocampal activity predicts memory accuracy

A Schematic of fMRI contrast. We predicted to see increased cross-stimulus suppression (XSS) on trials that were preceded by an associated stimulus, relative to control trials where a stimulus was preceded by a stimulus at least 4

links away on the associative ring (Figure 4.1C). The difference in BOLD signal between control and XSS trials was used as a neural index for coactivation between associated stimuli (memory XSS). **B** T-statistic map for coactivation between paired stimuli indexed using fMRI XSS. Across all participants, we observed evidence for coactivation between paired stimuli in hippocampus. (t_{40}=3.98, p=0.036, FWE peak-corrected using SVC; MNI coordinates). **C** Within a hippocampal ROI defined from the contrast in *B*, extracted parameter estimates (mean +/- SEM) illustrating hippocampal representations for the learned associations in ATX and PLC. **D** T-statistic map for coactivation between paired stimuli in the ATX group. We observed significant coactivation between paired associates in hippocampus (t_{21}=4.21, p=0.021, FWE peak-corrected using SVC; MNI coordinates). **E** Across all participants, the concentration of GABA+ in LOC predicted the coactivation between paired stimuli in hippocampus (ROI: Supplementary Figure 5.4D) (Spearman correlation: r_{33}=0.600, p<0.001). T-statistic maps are thresholded at p < 0.001 uncorrected for visualisation purposes only.

4.4.6 Relating neural memory expression to behavioural distortions in the cognitive map

We next asked whether coactivation of paired associates measured using fMRI XSS immediately after learning (Figure 4.5B) predicts subsequent behavioural measures used to detect distortions in the cognitive map, i.e. overgeneralisation errors reported in Figure 4.2F. We reasoned that if associated stimuli show stronger coactivation (e.g. bird '2' and bird '1'), then participants should be more likely to overgeneralise when tested on their implicit knowledge for the relationship between a given bird and context stimulus. In other words, participants erroneously report that bird '1' is paired with sofa 'B' rather than sofa 'A'. In line with this prediction, in the ATX group coactivation between paired stimuli in hippocampus predicted overgeneralisation errors measured in behaviour (Figure 4.6A). We did not observe this effect in the PLC group (Figure 4.6B) and in the hippocampus a significant difference was observed between the two groups (Figure 6C). These results suggest that under elevated noradrenaline, coactivation between memories predicts the subsequent tendency to overgeneralise across these memories.

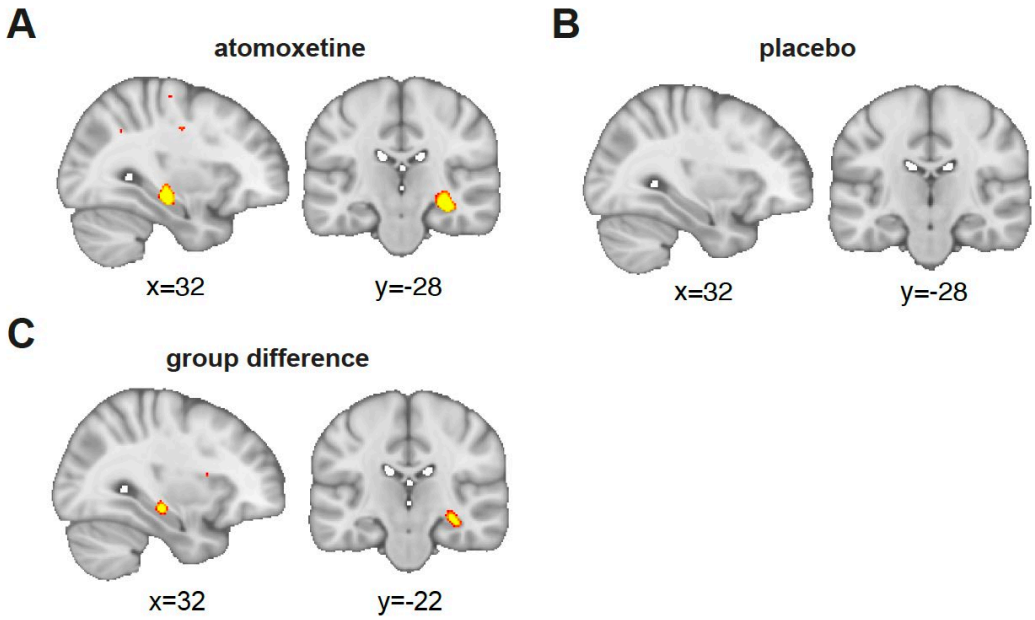


Figure 4.6 Memory expression in hippocampus following learning under atomoxetine predicts later behavioural distortions in the cognitive map
T-statistic maps for the regression effect between overgeneralisation errors and coactivation between paired stimuli (memory XSS index). **A** In the atomoxetine group, the strength of the XSS memory index effect in the right hippocampus predicted subsequent behavioural measures of overgeneralisation for implicit bird-context relationships tested on day 5. Across the whole brain, this effect was uniquely observed in the right hippocampus (t_{20}=6.01, p=0.012, whole-brain FWE peak-corrected; MNI coordinates). **B** In the placebo group there was no regression effect across the whole brain between the XSS memory index and overgeneralisation errors. **C** For the relationship described in *A-B*, we observed a significant group difference uniquely in right hippocampus (atomoxetine – placebo: hippocampus t_{39}=4.53, p=0.009, FWE peak-corrected using SVC; MNI coordinates). T-statistic maps are thresholded at $p < 0.001$ uncorrected for visualisation purposes only.

4.4.7 Elevated noradrenaline modulates neural overlap between stimuli and contexts

Having found evidence for representations of the explicitly learned associations between birds in hippocampus, we next turned to the implicit associations between stimuli (birds) and their contexts (sofas). We hypothesised that if inhibitory tone is reduced during learning, coactivation between stimuli should reflect a spread of association along the ring structure (Figure 4.1C, Figure 2C). Based on our behavioural finding that atomoxetine increases overgeneralisation of contextual cues (Figure 4.2G), we predicted to observe this neural increase in spread of association effect specifically for the implicit associations, namely between bird stimuli and contextual cues. Therefore, here we evaluated fMRI XSS effects in the scan blocks where bird stimuli and contextual cues were interleaved (Figure 4.3A). We predicted to see a group difference for trials where a bird stimulus was preceded by an indirectly associated contextual cue (sofa 'B' followed by bird '1') compared to control trials (sofa 'E' followed by bird '1'), which we denote as 'spread of association' XSS index.

In line with the prediction above, in hippocampus we observed a significant difference in our XSS index for 'spread of association' in the ATX group compared to PLC (Figure 4.7A).

Furthermore, across all participants, this neural spread of association effect in hippocampus was predicted by the concentration of GABA+ in LOC across all participants (Figure 4.7B). However, no main XSS effect was observed in the ATX, while a negative XSS effect was observed in PLC, indicating repetition enhancement (Supplementary Figure 4.6). Therefore, while the finding of a group difference suggests that elevated noradrenaline during learning induces a 'spread of association' effect in hippocampus, this difference appears to be driven by repetition enhancement effects in PLC.

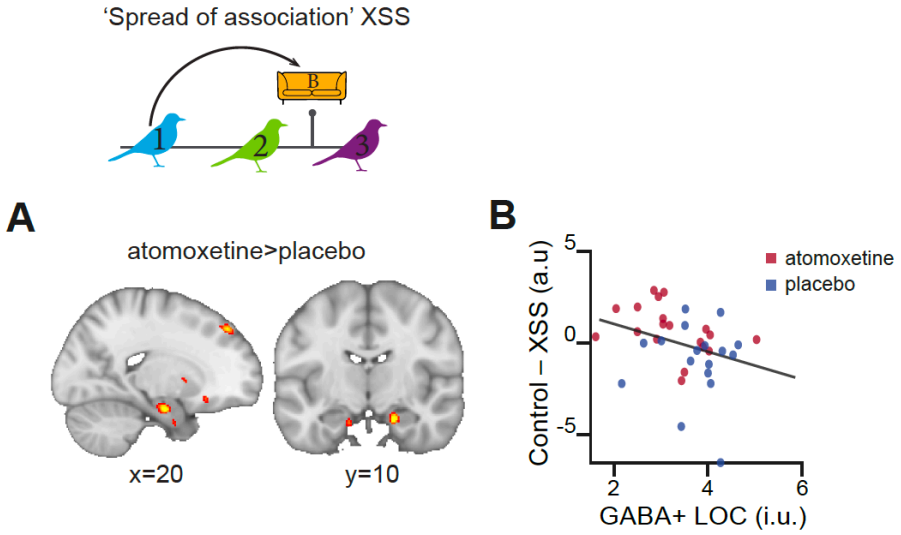


Figure 4.7 Atomoxetine induces hippocampal coactivation between stimuli and unassociated but relationally proximal contextual cues
A T-statistic map of the group difference (ATX – PLC) for the 'spread of association' XSS index, defined as the contrast between trials where the bird stimulus and preceding contextual cue were indirectly linked, versus trials where the bird stimulus and preceding contextual cue were at least 4 links separated along the associative ring (Figure 4.1C). In the ATX group we observed a relative increase in the 'spread of association' XSS index in hippocampus compared to placebo (ATX – PLC, hippocampus t_{41}=4.66, p=0.005, FWE peak-corrected using SVC; MNI coordinates). T-statistic map thresholded at $p < 0.001$ uncorrected for visualisation purposes only. **B** Across participants, the concentration of GABA+ in LOC negatively predicted the XSS effect in *A* (Spearman correlation: r_{34}=-0.398, p=0.017).

4.5 Discussion

Relationships between stimuli and events in the world are thought to be represented in the brain in a manner similar to a spatial map. However, these representations do not necessarily preserve the relationships in the world around us, but instead appear to contain behaviourally relevant distortions. Here, we investigated the circumstances under which such neural distortions might arise. Specifically, we asked whether the neuromodulator noradrenaline could facilitate plasticity by lowering inhibitory tone, introducing spurious mnemonic associations. We show that a single dose of a noradrenergic drug lowers cortical inhibition and increases neural and behavioural measures of distortions in the task-relevant cognitive map.

By combining a pharmacological intervention using the noradrenaline reuptake inhibitor atomoxetine with the learning of a cognitive map and acquisition of MRS and fMRI, we were able to relate behavioural readouts of memory distortions and neuronal excitability to the noradrenaline neuromodulatory system. We observed that a single dose of atomoxetine leads to an increase in overgeneralisation errors which are indicative of distortions in the task-relevant cognitive map. Furthermore, using MRS to measure the concentration of cortical GABA+ in LOC, we found that atomoxetine was associated with a relative decrease in cortical inhibition, which is in line with findings in rodents (Martins and Froemke, 2015). Using fMRI we quantified coactivation between neuronal representations of our task stimuli using cross-stimulus suppression, and find evidence for coactivation between representations of paired stimuli in hippocampus. Across all participants, the concentration of GABA+ in LOC predicted the coactivation between paired associates in hippocampus. Furthermore, in the atomoxetine group, the coactivation between paired associates in hippocampus predicted overgeneralisation errors observed in behaviour 4 days later.

The neuromodulator noradrenaline has been associated with signalling salience and modulating attention and neural gain (Aston-Jones and Cohen, 2005; Bouret and Sara, 2005; Eldar et al., 2013). Noradrenaline is primarily released from locus coeruleus (LC), which projects across the entire nervous system (Berridge, 2008; Fuxe et al., 2012). Activity in LC has been found to be related to pupil size (Gilzenrat et al., 2010; Joshi et al., 2016; Rajkowski et al., 1993), though see also (Megemont et al., 2022). For this reason pupil diameter has been used to index noradrenergic tone. Our finding that a single dose of a noradrenergic drug is associated with increased pupil dilation in response to a rare, surprising stimulus (Figure 4.3B) suggests that a single dose of atomoxetine indeed increased availability of noradrenaline. Our results are in line the suggestion that noradrenaline may be released to signal 'unexpected uncertainty' (Yu and Dayan, 2005). Here, the occurrence of a rare, unpredictable event would require updating of neural circuits through synaptic plasticity, which noradrenaline could facilitate through disinhibition. While we predicted that this physiological marker of noradrenergic tone could relate to our measures of inhibitory tone, or predict behavioural and neural markers of memory distortions, no such relationships were found in the current data. This is likely due to a low number of subjects with sufficiently good quality pupillometry data (see *Methods*).

At a physiological level, noradrenaline has been linked to decreasing tonic, ongoing inhibition (Martins and Froemke, 2015). In agreement with these findings, here we found that a single dose of atomoxetine is associated with a relative decrease in the concentration of MRS-derived GABA+ in LOC compared to placebo treatment. However, we only found this effect in LOC (Figure 4.4), and not in V1 (Supplementary Figure 4.3). Given the knowledge that LC projects across the entire

neocortex (Berridge, 2008; Fuxe et al., 2012), we predicted to observe a decrease in GABA+ in both LOC and V1. A potential explanation for the difference in effect we observed between LOC and V1 could be a difference in the availability of noradrenaline receptors between these two regions, given the heterogeneity and divergence of noradrenergic projections deriving from LC (Chandler and Waterhouse, 2012; Chandler et al., 2014). Indeed, it has been suggested that LC exerts differential control over different cortical areas (Kebschull et al., 2016).

Our finding that elevated noradrenergic tone in cortex lowers cortical inhibition in LOC indicates a shift in cortical E/I balance in favour of excitation (Froemke, 2015). We predicted that this shift towards cortical excitation facilitates plasticity, strengthening individual memories but also rendering acquired memories prone to distortion through the formation of stronger links between them. We designed our task such that all participants would have good memory for the explicitly learned associations (bird-bird), but limited memory for the implicit, contextual associations (bird sofa). This allowed for analysis of error trials on the implicit memory test. As a result of this task design, we found that atomoxetine had no effect on memory accuracy for the learned associations (Figure 4.2D) or associations between stimuli and contextual cues (Figure 4.2F), which otherwise might have been expected as memory is thought to be boosted by emotional arousal which is mediated by noradrenaline (Strange et al., 2003).

However, as we intended with our task design, participants made many errors on the implicit memory test. When analysing these error trials, those who received atomoxetine had an increased propensity to make overgeneralisation errors, where they preferentially link an indirectly associated contextual cue to the bird stimulus, rather than making a random error. We interpret these links between indirectly associated stimuli as distortions in the task-relevant cognitive map. The tendency to make these overgeneralisation errors was predicted by the degree of neural overlap between paired associates measured using fMRI immediately following learning (while still under the influence of atomoxetine). Combined with the result that inhibitory tone predicted coactivation between paired associates in hippocampus, these results suggest the possibility that reduced inhibitory tone during learning increases coactivation between paired stimuli, and that this coactivation when noradrenaline is elevated leads to the formation of distortions in the cognitive map (i.e. bird '1' evokes the representation of bird '2', which in turn evokes the representation of sofa 'B'). We note that, while we only measured this behavioural overgeneralisation or 'spread of association' effect between stimuli and contextual elements, it is possible that atomoxetine would induce a similar effect for those associations learned explicitly (i.e. bird-bird). However, we were unable to test this as by design our participants performed at ceiling on explicitly learned

associations on day 1 and we included only a low number of trials in the associative memory test on day 5 (Figure 4.2A, see *Methods*).

The hippocampal formation is thought to support the neural instantiation of cognitive maps (O'Keefe and Nadel, 1978). Here we used fMRI cross stimulus suppression (XSS) in hippocampus to index the neural overlap between paired associates as well as stimuli and contextual cues (Barron et al., 2016a; Grill-Spector et al., 2006). We show that the hippocampus represents the newly learned associations following learning, and that the strength of these representations is predictive of the formation of behavioural memory distortions in the ATX group. Furthermore, we show that atomoxetine modulates the neural overlap between stimuli and indirectly associated contexts. However, currently we do not have a clear interpretation of the differences between the XSS effects observed in the ATX versus PLC groups for the neural 'spread of association' effect. In the hippocampus we find that there is a significant group difference, but no main effect in the ATX group, and a repetition enhancement effect in PLC. Furthermore, we observed no XSS effects for either coactivation between paired stimuli in any cortical area, while these have been observed previously for newly learned associations between abstract stimuli immediately following learning (Barron et al., 2016a).

In this study, we found that a single dose of a noradrenaline reuptake inhibitor led to an increase in unwanted overgeneralisation between stimuli and contextual cues that were associated in an implicit manner. This behavioural result may shed light on the memory distortions observed in individuals following traumatic events, which are typically accompanied by release of neuromodulators involved in arousal, such as noradrenaline. In the extreme, learning in states of stress or high arousal may render a memory maladaptive as observed in Post-Traumatic Stress Disorder where levels of noradrenaline are higher than healthy controls and predict the severity of clinical symptoms (Kosten et al., 1987; Yehuda et al., 1998). Our results support the suggestion that these clinically observed maladaptive memories may be attributed to increased noradrenergic tone during memory acquisition, and suggest that noradrenaline may induce these distortions in cognitive maps by modulating inhibitory tone.

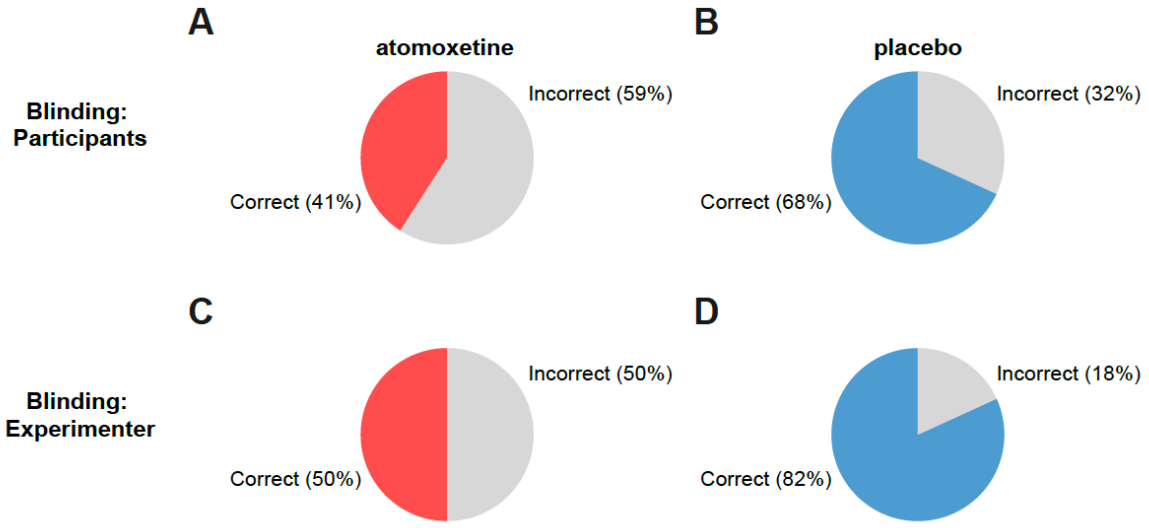
Finally, we note that while atomoxetine is a noradrenaline reuptake inhibitor, the noradrenergic and dopaminergic neuromodulatory systems are not independent of each other, particularly in prefrontal cortex where dopamine reuptake is regulated by noradrenaline transporters (Bymaster et al., 2002). Indeed, atomoxetine has been found to increase both extracellular levels of noradrenaline as well as dopamine in rodent studies (Bymaster et al., 2002; Swanson et al., 2006).

Therefore we cannot rule out the possibility that some of the effects measured in this study are in part due to changes in prefrontal dopamine availability.

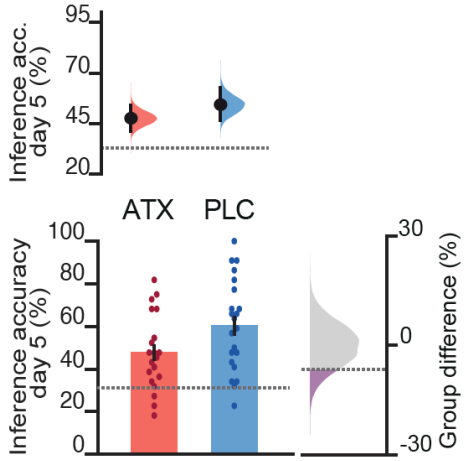
4.6 Conclusion

In this chapter I aimed to investigate the effect of a perturbation in E/I balance during learning on the subsequent memories formed. I showed that a single dose of a noradrenergic drug during learning lowers GABA in neocortex, and that this manipulation results in distortions in behavioural readouts of implicit memory. These results suggest that neuromodulators may regulate the formation of cognitive maps during new learning by setting the trade-off between stability and flexibility.

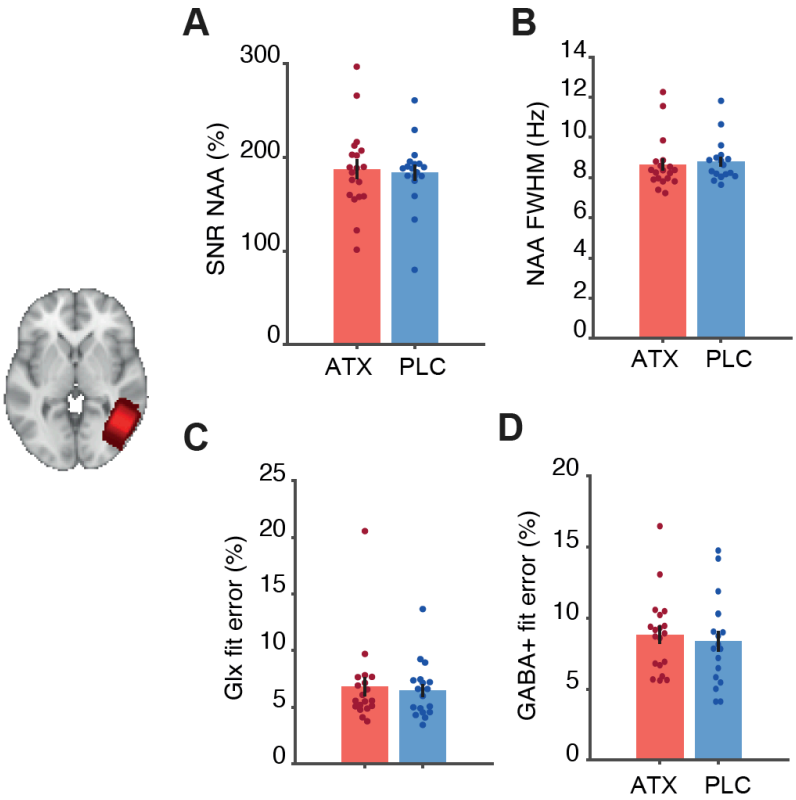
4.7 Supplementary figures



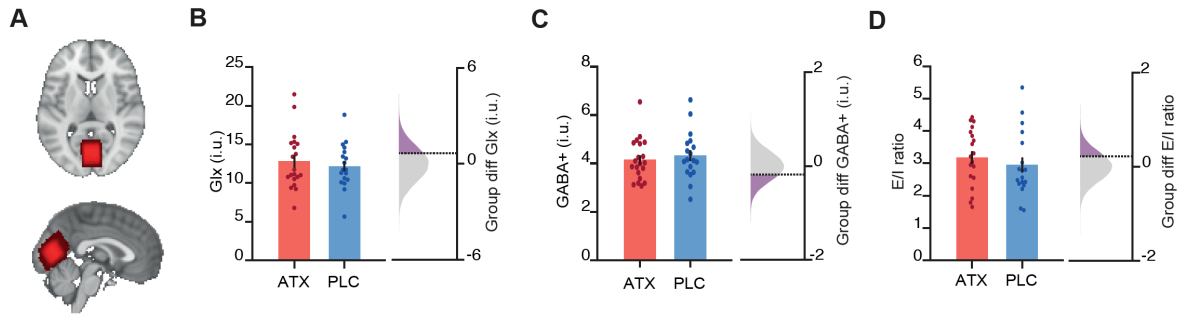
Supplementary Figure 4.1 Double-blinding results
To assess the effectiveness of the double-blinding procedure, after completion of the first study day the participants and the experimenter were asked whether they thought they/the participant has received drug or placebo. **A** In the atomoxetine group, 41% of participants correctly guessed that they had received atomoxetine. **B** In the group, 68% of participants correctly guessed that they had received placebo. **C** The experimenter correctly guessed for 50% of participants that they had received atomoxetine. **D** The experimenter correctly guessed for 82% of participants that they had received placebo.



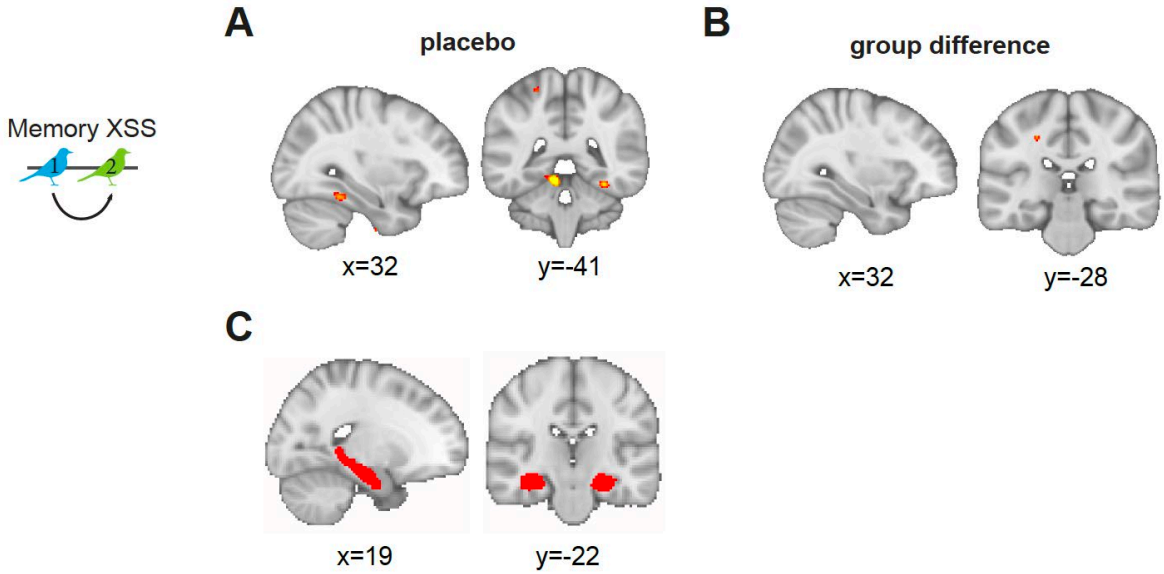
Supplementary Figure 4.2 Inference test day 5 performance
On day 5 participants repeated the associative inference test (Figure 4.1F). There was no significant difference between inference task performance between the groups (ATX – PLC: permutation test: p=0.124).



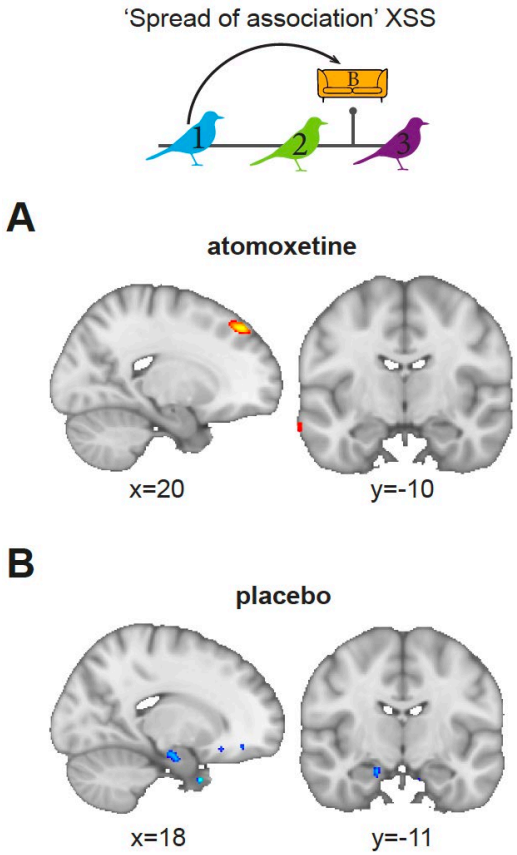
Supplementary Figure 4.3 MRS quality measures for LOC do not differ between groups
A The average signal-to-noise ratio (SNR) of N-acetylaspartate (NAA), a reference metabolite, as determined by Gannet was for 187.1% (SEM 10.24%) the ATX and 183.2% (SEM 8.39%) for PLC group. There was no significant difference between groups (permutation test: p= 0.388). **B** The average full-width-at-half-max (FWHM) of NAA as determined by Gannet was for 8.638 Hz (SEM 0.296) the ATX and 8.777 Hz (SEM 0.240) for PLC group. There was no significant difference between groups (permutation test: p= 0.363). **C** The average fit error for Glx as determined by Gannet was for 6.820% (SEM 0.835) the ATX and 6.447% (SEM 0.563) for PLC group. There was no significant difference between groups (permutation test: p= 0.379). **D** The average fit error for GABA+ as determined by Gannet was for 8.814% (SEM 0.632) the ATX and 8.344% (SEM 0.710) for PLC group. There was no significant difference between groups (permutation test: p= 0.312).



Supplementary Figure 4.4 Noradrenaline does not induce changes in Glx and GABA+ in V1
A Anatomical location of the 2 x 2 x 2 cm^3 MRS VOI, positioned in V1. Cumulative map across 38 participants. **B-D** Left: metabolite concentrations and E/I ratio (mean +/- SEM). Right: null distribution of the group differences generated by permuting subject labels, dotted line: true group difference. **B** No significant group difference was observed in the concentration of Glx (Glx ATX – PLC: permutation test p=0.267). **C** No significant group difference was observed in the concentration of GABA+ (GABA+ ATX – PLC: permutation test p= 0.269). **D** There was no significant group difference in E/I ratio, defined as the ratio of Glx:GABA+ (E/I ratio ATX – PLC: permutation test p=0.233).



Supplementary Figure 4.5 Coactivation between paired stimuli per group and group differences
A-B T-statistic maps for the representation of the learned associations (memory XSS index) for PLC (A) and for the group difference (B). **A** In the PLC group, the memory XSS cluster in parahippocampus did not survive whole-brain correction (t_{19}=3.95, p=0.907, whole-volume FWE peak-corrected; MNI coordinates). **B** For the group difference (ATX – PLC), no clusters survived whole-brain or small volume correction. **C** Hippocampal region of interest (ROI), used for the analysis in Figure 5E. MNI coordinates. T-statistic maps are thresholded at $p < 0.001$ uncorrected for visualisation purposes only.



Supplementary Figure 4.6 'Spread of association' XSS per group
T-statistic maps per group for the 'spread of association' XSS index, defined as the contrast between trials where the probe and preceding contextual stimulus were indirectly linked, versus trials where the probe and preceding contextual stimulus were at least 4 links away on the associative ring (Figure 4.1C). **A** In the atomoxetine group, no clusters survived whole-brain or small volume correction. **B** Blue: negative repetition suppression effect (otherwise termed repetition enhancement). In the placebo group, we observed evidence for repetition enhancement between stimuli and indirectly linked contextual stimuli in hippocampus (t_{21}=3.81, p=0.047, FWE peak-corrected using SVC; MNI coordinates). T-statistic maps are thresholded at $p < 0.001$ uncorrected for visualisation purposes only.

CHAPTER 5

THE HIPPOCAMPUS AND NEOCORTICAL INHIBITORY ENGRAMS PROTECT AGAINST MEMORY INTERFERENCE

The previous chapter showed that a decrease in inhibitory tone during learning resulted in a distorted task-relevant cognitive map that is more prone to neural and behavioural overgeneralisation. Here I asked what mechanisms normally prevent overgeneralisation or interference between overlapping memories. In line with the previous chapter, I predicted that hippocampus would represent the underlying relational structure of the learned information, encoding relevant contextual information. Furthermore, I predicted that cortical inhibition would normally prevent coactivation of stored overlapping memories. To test these predictions, I measured representational overlap between two sets of overlapping memories in hippocampus, and found that overlapping memories are separated using relational information. To reveal the role of inhibition in neocortex I lowered cortical GABA by applying transcranial Direct Current Stimulation. The drop in GABA was associated with an increase in interference between the overlapping memories. These results reveal two distinct neural mechanisms, hippocampal relational codes and cortical inhibition, that prevent interference between overlapping memories.

* This chapter is published in the following article:

Koolschijn RS, Emir UE, Pantelides AC, Nilli H, Behrens TEJ, Barron HC (2019). The hippocampus and neocortical inhibitory engrams protect against memory interference. *Neuron* 101: 1–14

5.1 Abstract

Our experiences often overlap with each other, yet we are able to selectively recall individual memories to guide decisions and future actions. The neural mechanisms that support such precise memory recall remain unclear. Here, using ultra-high field 7T MRI we reveal two distinct mechanisms that protect memories from interference. The first mechanism involves the hippocampus, where the BOLD signal predicts behavioural measures of memory interference, and representations of context-dependent memories are pattern separated according to their relational overlap. The second mechanism involves neocortical inhibition: when we reduce the concentration of neocortical GABA using trans-cranial direct current stimulation (tDCS) neocortical memory interference increases in proportion to the reduction in GABA, which in turn predicts behavioural performance. These findings suggest that memory interference is mediated by both the hippocampus and neocortex, where the hippocampus separates overlapping but context-dependent memories using relational information, while neocortical inhibition prevents unwanted co-activation between overlapping memories.

5.2 Introduction

Our decisions and actions are often guided by past experiences that overlap with one another in their content or sensory information. To ensure that interference between related or overlapping experiences is minimised, a stable memory storage system is critical. However, the precise physiological mechanism that supports stable memory storage in the absence of memory interference remains unclear.

One way to minimise memory interference is to separate stored information using contextual representations (McClelland et al., 1995; Norman and O'Reilly, 2003; Shapiro and Olton, 1994). Behavioural data in humans provides supporting evidence for this mechanism, as contextual cues help mitigate memory interference between two lists of paired associates (Bilodeau and Schlosberg, 1951). At the neural level, anticorrelated firing patterns for opposing contexts can be observed in the hippocampal output regions (Butterly et al., 2012; McKenzie et al., 2014). These contextual representations likely reflect the natural consequence of pattern separation, a competitive mechanism supported by the architecture of the hippocampus that orthogonalises representations of overlapping memories (Yassa and Stark, 2011). However, it remains unclear how information within contextual representations is organised. One possibility is that contextual information is organised in a manner that reflects the relational or configural structure of memory elements

(Cohen and Eichenbaum, 1993; Sutherland and Rudy, 1989). Consistent with this suggestion, here we hypothesise that the hippocampus helps protect against memory interference by separating elements of a partially overlapping memory according to their relational similarity. Thus, elements that have different relational positions across two overlapping but context-dependent memories are maximally separated.

In addition, an alternative way to protect stored memories from interference involves using inhibition. Following learning, new information is thought to be stored in the brain via modification in the strength of excitatory connections (Hebb, 1949; Nabavi et al., 2014; Song and Abbott, 2001). In turn, these newly modified excitatory connections are opposed by equivalent changes in the strength of inhibitory connections (Barron et al., 2016a; Froemke et al., 2007; Vallentin et al., 2016; Vogels et al., 2011). This allows excitatory-inhibitory (E/I) balance to be maintained despite new learning (Froemke et al., 2007; Haider et al., 2006; Okun and Lampl, 2008; Wehr and Zador, 2003), and ensures that memories lie dormant unless E/I balance is disturbed (Barron et al., 2016a; Jacobs and Donoghue, 1991; Vallentin et al., 2016). Here we hypothesise that the inhibitory component of a memory, otherwise termed the inhibitory engram (Barron et al., 2017), protects memories from interference by preventing run-away excitation.

Consistent with this hypothesis, context-dependent behaviour in rodents is accompanied by modulation of neocortical interneurons (Kuchibhotla et al., 2017), while in humans an increase in neocortical GABA relative to glutamate accompanies overlearning, a process known to protect memories from interference (Shibata et al., 2017). Clinical investigations also support a key role for inhibitory regulation of memory expression, as impaired GABAergic regulation can readily account for delusions and hallucinations reported in schizophrenia (Vogels and Abbott, 2007; Yizhar et al., 2011). Thus, by gating memory expression (Barron et al., 2017, 2016a; Vogels and Abbott, 2009), inhibitory engrams may play a critical role in preventing unwanted interference between overlapping memories.

Here we investigate the role of both the hippocampus and neocortical inhibition in protecting against memory interference. First, we test the hypothesis that contextual representations in the hippocampus are organised using a relational code, thus separating competing memories according to behaviourally relevant information. Second, we test the hypothesis that neocortical inhibition protects overlapping memories from interference.

To this end, we designed a task that required participants to encode two overlapping but context-dependent memories across two consecutive days. On the third day, interference between the two memories was measured using ultra-high field 7T Magnetic Resonance Imaging (MRI). In

the hippocampus, we observed an increase in Blood-Oxygen-Level Dependent (BOLD) signal during opportunities for memory interference which predicted subsequent behavioural performance. In addition, representations of stimuli that had different relational positions across the two overlapping but context-dependent memories were maximally separated. Then, to investigate the role of neocortical inhibition in protecting memories from interference, halfway through the scan we manipulated the concentration of neocortical GABA using brain stimulation, and re-assessed evidence for memory interference. The drop in neocortical GABA induced by brain stimulation predicted an increase in neocortical memory interference, which in turn predicted deficits in behavioural performance. Together these results suggest that memory interference is mediated by two distinct mechanisms: a hippocampal mechanism where contextual representations are organised according to behaviourally relevant relationships, and a neocortical mechanism where inhibition protects overlapping memories from unwanted co-activation.

5.3 Methods

5.3.1 Participants

91 healthy volunteers participated in the study (group 1: "MRI" with tDCS, n = 30, mean age of 24.0, 17 females; group 2: "tDCS", n = 20, mean age of 21.9, 18 females; group 3: "Sham", n = 20, mean age of 24.1, 9 females; group 4: "Behav", n =21, mean age of 23.2, 11 females). All experiments were approved by the Oxford University ethics committee (reference number ref R43594/RE001). All participants gave informed written consent.

In group 1, three participants dropped out after the first day as they were not able to achieve the day 1 training criteria (see below). In group 1, one participant was excluded due to a fault on the scanner which prevented data acquisition in the second half of the scan session. In group 4, one participant was excluded after revealing that they had an arteriovenous malformation in the cerebellum.

5.3.2 Behavioural training

All behavioural tasks were coded in Matlab 2016b using Psychtoolbox (version 3.0.13). Seven different stimuli were presented to the participant, referred to as 1:7. Stimuli were rotationally invariant gray shapes (Figure 5.1A), which were observed in one of four possible rotations, with each rotation separated by 90° (as described in Barron et al., 2016a). Learned associations between these rotating shapes are known to be represented in a localised and superficial region of neocortex

(Barron et al., 2016a), thus providing a suitable target for anodal tDCS (see below). The experiment was conducted across three days (Figure 5.1F). On the first day participants performed a training task, see below, to learn seven bidirectional associations between the seven stimuli. The set of associations could be arranged in a ring structure (Figure 5.1B), where each stimulus was associated with two other stimuli (1 with 2, 2 with 3, etc. and 7 with 1). Participants were not explicitly made aware of the ring structure. Stimulus allocation within the ring structure was randomised across participants using Matlab's random number generator.

The training task included two phases. During phase 1 of the training task participants were passively exposed to seven pairs of associated stimuli. On each trial of phase 1, a pair of associated stimuli was presented against a background colour (blue or yellow, depending on the training day) for 3 s duration (see Supplementary Figure 5.1A-B). The stimulus presented on the left hand-side was randomised on each trial. Each pair of stimuli was presented four times in total, once for each of the four possible orientations of each stimulus. Across trials, the order in which pairs of stimuli were presented was randomised. The background colour was different on day 1 and day 2, thus providing a contextual cue for the learned associations. Participants were allowed to repeat phase 1 of the training task before each phase 2 block if they wished to do so.

During phase 2 of the training task participants performed an active task, involving a three-alternative forced-choice task (Figure 5.1D-E). On each trial of the three-alternative forced-choice task, one of the seven stimuli was shown as a probe stimulus for 1 s before three choice stimuli were presented in randomised positions across the screen. As in phase 1, stimuli were presented against a background colour (blue or yellow, depending on the training day), used to provide a contextual cue. The three choice stimuli included one stimulus to which the probe stimulus was associated, and two stimuli to which the probe stimulus was not associated. Participants were instructed to select the correctly paired stimulus as fast as possible, without compromising their accuracy, using the appropriate keyboard button 'b', 'n, or 'm'. If participants failed to make a response within 3 s they received an on-screen message indicating that they were too slow. Participants received feedback for each choice, where the probe stimulus together with the correctly paired choice stimulus was presented for 1.5 seconds. For each correct response, participants were assigned 50p. Each task block included 100 trials in total, and each pair of associated stimuli was presented at least 14 times. Across trials, the rotation of the presented stimuli and the trial order were randomised. At the end of each task block three percent of trials were randomly selected and participants received the sum total reward from these trials.

On the second day, participants again learned seven associations between the seven stimuli, however the position of the stimuli within the implicit ring structure was altered relative to the first day. In particular, stimuli '3' and '6' were switched, resulting in four new associations and three associations that remained the same across days (Figure 5.1C). To indicate this change in the implicit associative structure, the background colour of the screen (blue or yellow) was changed from day 1 to day 2. The colour assigned to day 1 and 2 was randomised across participants. To learn the new arrangement of stimuli, participants underwent both phase 1 and 2 of the training task again, thus following the same protocol as used on day 1 (Supplementary Figure 5.1B, Figure 5.1E).

On both day 1 and 2, the criterion for stopping the phase 2 training task was as follows: participants were required to complete at least five blocks and reach 90% accuracy on all of the seven associations. If after five blocks of phase 2 of the training task participants did not reach the criterion of 90% accuracy on all of the seven associations, then they were required to continue completing phase 2 task blocks until this criterion was met (Supplementary Figure 5.1F). By ensuring that participants completed at least five blocks of phase 2 of the training task on both day 1 and 2, our experimental protocol minimised differences in the number of phase 2 trials completed on day 1 compared to day 2, and minimised differences in the number of phase 2 trials completed across participants. Across all 4 experimental groups, 3 out of 91 participants dropped out of the experiment after not reaching 90% accuracy on all seven associations despite completing multiple phase 2 training task blocks. Data from these participants was not included in any analyses. As some participants got tired during the final training blocks of the phase 3 training task, performance during the final and/or penultimate task block was sometimes compromised. For this reason, 'learning accuracy' on both day 1 and 2 was estimated as the average performance across trials on each participant's highest performing task block (Supplementary Figure 5.1C-D).

On the third day of the experiment, participants were required to perform the fMRI scan task (see below, Figure 5.1G). The fMRI scan task was performed inside the scanner for group 1, but outside the scanner for groups 2-4. Immediately after exiting the scanner (group 1), or immediately after the scan task (groups 2-4), participants were given a surprise memory test designed to assess participants' memory for the associations learned on both day 1 and day 2 (Figure 5.2D). The memory test involved a variant of the three-alternative forced-choice task used during training on day 1 and 2 (Figure 5.1D-E). However, unlike the training task, the background colour switched randomly between trials to indicate either the day 1 or day 2 context, and the task was presented in the absence of feedback. Given the probe stimulus and the background colour, participants

were instructed to select the correct associated stimulus. The memory test constituted 100 trials, with half presented on the yellow background and half on the blue background.

5.3.3 fMRI scan task

The fMRI scan task involved participants viewing the seven visual stimuli used in the training task (1:7), presented via a computer monitor, which for group 1 was then projected onto a screen inside the scanner bore. On each trial two stimuli were presented consecutively for 800 ms each, with an inter-stimulus interval of 300 ms (Figure 5.1G). The inter-trial interval was selected from a truncated gamma distribution with mean of 2.9 s, minimum of 1.5 s and maximum of 9.7 s. To control for potential confounding effects of expectation suppression (Summerfield et al., 2008), all stimuli, all possible pairs of non-repeating stimuli, and all possible rotations of each stimulus were presented equally often in a fully randomised order. Participants were required to perform a task incidental to the contrast of interest which involved identifying whether the presented stimuli were familiar or "oddball." Oddball stimuli, defined as stimuli that did not belong to the training set 1 to 7, were randomly inserted into 7% of trials. Participants were instructed to press a button on an MR compatible button box using their right index finger when they identified "oddball" stimuli but not if both stimuli on the trial were familiar. No feedback was given. Each task block lasted twenty minutes and included 196 trials, with each stimulus presented 52 times, with 13 examples of each of the four possible rotations per stimulus. Within each block, each pair of non-repeating stimuli was presented 4 times in each context, while each pair of repeating stimuli was presented 2 times in each context. Each participant performed two task blocks.

5.3.4 fMRI imaging protocol

Participants in group 1 completed the scan task within a 7 Tesla Magnetom MRI scanner (Siemens) with 1-channel transmit and a 32-channel phased-array head coil (Nova Medical, USA) at the Wellcome Centre for Integrative Neuroimaging Centre (University of Oxford). Current 7T radio-frequency (RF) coil designs suffer from B1 inhomogeneity effects which were pronounced in the right temporal lobe. To overcome this, we positioned two barium titanate dielectric pads (4:1 ratio of BaTiO3:D2O, with a relative permittivity of around ~300, and size 110 × 110 × 5 mm3) over the right temporal lobe in all 7T scanning sessions, causing a "hotspot" in the RF distribution at the expense of distal regions (Brink and Webb, 2014; Teeuwisse et al., 2012). The tDCS electrode was situated between the dielectric pad and the head.

To acquire fMRI data a multiband echo planar imaging (EPI) sequence was used to acquire 50 1.5 mm thick transverse slices with 1.5 mm gap, in-plane resolution of 1.5 × 1.5 mm^2, repetition time (TR) = 1.512 s, echo time (TE) = 20 ms, flip angle = 85°, field of view 192 mm, and acceleration factor of two. To increase SNR in brain regions for which we had strong prior hypotheses, we restricted the fMRI sequence to a partial volume, thus increasing the number of measurements acquired. The partial volume covered occipital and temporal cortices (see Supplementary Figure 5.2A-C) and in each session 644–723 volumes were collected (~20 min). For each participant, a T1-weighted structural image was acquired to correct for geometric distortions and co-registration between EPIs, consisting of 176 0.7 mm axial slices, in-plane resolution of 0.7 × 0.7 mm^2, TR = 2.2 s, TE = 2.96 ms, and field of view = 224 mm. A field map with dual echo-time images was also acquired (TE1 = 4.08 ms, TE2 = 5.1 ms, whole-brain coverage, voxel size 2 × 2 × 2 mm^3).

5.3.5 Magnetic Resonance Spectroscopy

For participants in group 1, during the scan session MRS data was acquired as described in (Barron et al., 2016a). B^0 shimming was performed in a two-step process. First, GRE-SHIM (field of view, 384 × 384 mm^2; TR = 600 ms; TE1/2 = 2.04/4.08 ms; slice thickness 4 mm; flip angle 15°; slices 64; scan time 45 s) was used to determine the optimal first- and second-order shim currents. The second step involved only fine adjustment of first-order shims using FASTMAP (Gruetter and Tkác, 2000). The modified semi-LASER sequence, previously shown to have minimal chemical shift displacement error (CSDE), was used with TE = 36 ms, TR = 5–6 s to acquire MRS measurements in a 2 × 2 × 2 cm^3 volume of interest (VOI), positioned next to the tDCS electrode (Figure 5.4C) (Öz and Tkáč, 2011).

For each MRS measurement between 65 and 130 scan averages were collected, giving a total acquisition time of around 10 min. Three measurements were acquired for each participant, before and during tDCS, and after the second task block (Figure 5.1F). Metabolites were quantified using LCModel (for example spectra: Supplementary Figure 5.3A-B) (Provencher, 2001, 1993). The model spectra of alanine (Ala), aspartate (Asp), ascorbate/vitamin C (Asc), glycerophosphocholine (GPC), phosphocholine (PCho), creatine (Cr), phosphocreatine (PCr), GABA, glucose (Glc), glutamine (Gln), glutamate (Glu), glutathione (GSH), myo-inositol (myo-Ins), Lactate, N-acetylaspartate (NAA), N-acetylaspartylglutamate (NAAG), phosphoethanolamine (PE), scyllo-inositol (scyllo-Ins) and taurine (Tau) were generated based on previously reported chemical shifts

and coupling constants by VeSPA Project (Versatile Simulation, Pulses and Analysis) (Govindaraju et al., 2000; Tkáč, 2008).

The unsuppressed water signal acquired from the VOI was used to remove residual eddy current effects and to reconstruct the phased array spectra (Natt et al., 2005). To improve comparability across spectra, the water component of the spectra was then removed before single scan spectra were summed from 32 channels, corrected for frequency and phase variations induced by participants' motion, and then summed. LCModel analysis was performed on all spectra within the chemical shift range 0.5 to 4.2 ppm (Provencher, 1993).

Reliable LCModel fits were achieved in 20 of the 26 participants and metabolite concentration relative to total Creatine concentration were estimated, relative to unsuppressed water spectrum acquired from the same VOI. In the remaining 6 participants the relative GABA quantification was either unreliable or inestimable due to lipid contamination and broader linewidth. The lipid contamination could be observed directly in the spectral range 1.9-0.5 ppm (Supplementary Figure 5.3A-B). The broader linewidth, quantified using Full-Width at Half Maximum (FWHM), was significantly higher in these six participants relative to the 20 participants included for analysis (two sample t-test: $t24=3.73$, $p=0.001$, Supplementary Figure 5.3C). Participants with inestimable GABA were excluded from all data analyses that included MRS data.

All measured metabolites included in the analysis had Cramér–Rao lower bound (CRLB) values $\leq 50\%$ (Bednařík et al., 2015). Relative to baseline concentrations ('Before tDCS'), the change in relative GABA (Figure 5.4E), relative glutamate (Figure 5.4F), and other metabolite concentrations (Supplementary Table 5.1) were compared across conditions using a two-tailed paired t-test where the direction of the effect was unknown and a one-tailed paired t-test in instances where the direction of the effect was predicted from previous data (i.e. for the change in relative GABA). Thus, all t-tests were performed using within-subject comparisons.

5.3.6 Transcranial Direct Current Stimulation

Immediately before and during Block 2 of the scan task, participants in groups 1, 2 and 3 received tDCS using a DC-Stimulator (Eldith) which delivered a 1 mA current to the brain. For group 1, the current was delivered while participants were inside the 7T MRI scanner. For groups 2 and 3, the current was delivered outsider the scanner using a double-blind procedure (see below). To ensure that the tDCS was suitable for use in the 7T scanner, we used two 5×7 cm^2 MRI compatible electrodes (Easycap) fitted with 5 kOhm resistors to minimise the risk of heating or eddy current induction. Using high-chloride EEG electrode gel (Easycap) as a conducting paste,

the anodal electrode was placed on the scalp above the region of right temporal cortex previously identified as encoding the association between paired shapes (Figure 5.4A-B), approximately at the 10–20 T6 node location. The cathodal electrode was placed over the contralateral supraorbital ridge (Figure 5.4A, Supplementary Figure 5.5). For participants in group 1, a cod-liver oil capsule was taped to the anodal electrode, immediately underneath the resistor, to make the electrode MR-visible and allow for its location to be mapped onto the anatomical brain surface (Figure 5.4C). For all participants, the impedance of tDCS was checked prior to stimulation. In group 1, this impedance check was performed before participants entered the scanner and again once the participant was lying inside the bore of the magnet with extension leads connected to the stimulator. For participants in group 1 and 2, tDCS was delivered using a 10 s ramp-up of the current, which was then held at 1 mA current for a total of 20 min, before a 10s ramp-down. For participants in group 3, sham stimulation involved mimicking the prickling sensation of stimulation using a 10s ramp-up of current, turning stimulation off for 20 minutes and then repeating the 10s ramp-up. For participants in all groups 1-3, the stimulation protocol commenced 10 min prior to the start of the second fMRI scan task (Figure 5.1F). At the end of the experiment, participants in groups 2 and 3 were debriefed: they were informed that they may have received sham stimulation and were asked to declare whether they believed they had received real or sham stimulation. The blinded researcher (R.K., see below) also declared whether they believed the participants had received real or sham stimulation.

5.3.7 Double-blind procedure for anodal/sham tDCS

Participants in groups 2 and 3 were first recruited, before being randomly assigned to the anodal (group 2) or sham (group 3) stimulation condition using a random number generator. Randomisation was performed by a researcher (H.B.) who was not involved in recruitment. Behavioural training, electrode placement, the scan tasks, surprise memory test and debrief were carried out by a researcher blind to the stimulation condition (R.K.). tDCS was delivered by a researcher who was aware of the stimulation conditions and who was not involved in any of the behavioural training or assessment (H.B.). Analysis of participants' responses during the debrief indicated that 55% of participants in group 2 ('tDCS') and 75% of participants in group 3 ('sham') believed they received anodal tDCS stimulation. The blinded researcher believed that 55% of participants in group 2 ('tDCS') and 45% of participants in group 3 ('sham') received anodal tDCS stimulation.

5.3.8 fMRI data analysis

For all MRI data sets obtained from participants in group 1, pre-processing was carried out using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Two participants were excluded from the fMRI analysis due to poor performance on the fMRI scan task (<80% accuracy on one or more of the two task blocks), suggesting that they may have fallen asleep during the task. For the remaining 24 participants images were corrected for signal bias, realigned to the first volume, corrected for distortion using field maps, normalised to a standard EPI template and smoothed using an 8-mm full-width at half maximum Gaussian kernel. To remove low frequency noise from the preprocessed data, a high-pass filter was applied to the data using SPM12's default settings. For each participant and for each scanning block, the resulting fMRI data was analysed in an event-related manner using two different general linear models (GLMs), one designed for univariate analyses and a second designed for multivariate analyses. In both GLMs explanatory variables used a delta function to indicate the onset of a trial and were then convolved with the hemodynamic response function.

The first GLM, used to analyse univariate BOLD effects, was applied to data from each of the two scan task blocks separately, and to data from both scan task blocks together. In the design, a total of 46 different explanatory variables were included per block. 42 of these explanatory variables were included to account for each possible pair of visual stimuli ('1' and '2', '1' and '3' etc.) in each of the two background contexts (i.e. memory 1 or memory 2), regardless of the order in which the two stimuli were presented within the pair. An additional 4 explanatory variables were used to model trials that included repeating stimuli or trials were 'odd-ball' stimuli had been presented in each of the two background contexts (i.e. memory 1 or memory 2). Finally, for each task block an additional 6 scan-to-scan motion parameters produced during realignment were included in the GLM as additional nuisance explanatory variables to account for motion-related artefacts.

Using the output of this first GLM for the univariate analysis, the following three principal contrasts of interest were assessed. First, to measure cross-stimulus adaptation as an index for expression of directly associated stimuli (Supplementary Figure 5.4), the contrast of interest involved comparing fMRI BOLD signal for trials with pairs of stimuli separated by more than one link across both memories ('unassociated'; i.e. memory 1 and memory 2 links 2-7, 5-7, 2-4, 1-5, 1-4, 2-5, 4-7, 3-6, 1-6, 1-3) with fMRI BOLD signal for trials with pairs of stimuli separated by one link in both memories ('associated'; i.e. memory 1 and memory 2 links 1-2,4-5,7-1). Second, to measure cross-stimulus adaptation as an index for memory interference (Figure 5.5, Figure 5.6A-

B), the contrast of interest involved comparing fMRI BOLD signal for pairs of stimuli separated by more than one link across both memories ('unassociated'; i.e. memory 1 and memory 2 links 2-7, 5-7, 2-4, 1-5, 1-4, 2-5, 4-7, 3-6, 1-6, 1-3) with fMRI BOLD signal for pairs of stimuli separated by more than one link in the current context, but only one link in the alternative context ('hidden'; i.e. memory 1: links 3-5, 4-6, 2-6, 3-7; memory 2: links 3-4, 5-6, 2-3, 6-7). Third, to measure the BOLD response to trials where there was an opportunity for memory interference (Figure 5.2A-C, 2H, 6C), the contrast of interest involved comparing fMRI BOLD signal for pairs of stimuli that shared the same topological relationship across the two memories ('stable', i.e. links that did not include stimuli 3 or 6) with fMRI BOLD signal for pairs of stimuli that had a different topological relationship across the two memories ('unstable'; i.e. links that included stimuli 3 or 6).

5.3.9 fMRI statistics and ROI specification

From the first GLM, the contrast images of all participants were entered into a second-level random effects analysis. To test for fMRI cross-stimulus suppression effects in an unbiased fashion, parameter estimates obtained from the relevant GLM were extracted from an independent region of interest (ROI) (see below). Paired t-tests were used to assess differences in the main effect between sessions. When testing evidence for replication of our previous findings (Barron et al., 2016a) a one-tailed test was used. In all other instances, two-tailed tests were used. The significance level was set to $p<0.05$.

To assess fMRI cross-stimulus suppression effects in the neocortex, three ROIs were defined. To assess evidence for replication of previously published results (Supplementary Figure 5.4), an independent ROI was defined using the previously published dataset, after thresholding the contrast of interest at $p<0.01$ uncorrected (Figure 5.4B, Supplementary Figure 5.4B) (Barron et al., 2016a). To assess evidence for memory interference, an ROI was defined from the peak average location of the anodal tDCS electrode. This was estimated using the T1 scan to identify the location of the cod-liver oil capsule taped immediately underneath the resistor of the anodal electrode. For each participant, the ventral-dorsal coordinate was taken from the upper edge of the cod-liver capsule. The medial-lateral coordinate was projected 20mm from the lateral surface, consistent with the peak medial-lateral coordinate of previously published cross-stimulus suppression induced by application of tDCS (Figure 5.4B) (Barron et al., 2016a). For each participant, an 8 mm sphere was then drawn around the identified coordinate, and the sphere was warped to a standard EPI template. Across individuals the average ROI was calculated (Figure 5.4C). To perform SVC for multiple comparisons, a 10mm sphere was drawn around the peak of the group average tDCS

electrode location (Figure 5.4D) and statistical significance assessed using peak-level FWE correction at $p<0.05$. Capitalising on variance across participants within this larger ROI, the extracted fMRI cross-stimulus suppression measures were correlated with changes in GABA and behaviour (Figure 5.5B-D).

To assess changes in BOLD signal in the hippocampus (Figure 5.2), an anatomical hippocampal mask was used to perform SVC for multiple comparisons (Figure 5.2G), with peak-level FWE correction at $p<0.05$.

5.3.10 Representational Similarity Analysis

The second GLM was used to assess multivariate effects. In this GLM each trial was modelled as a unique explanatory variable. All trials across Block 1 and 2 were included. In addition, 6 scan-to-scan motion parameters produced during realignment were included in the GLM as additional nuisance explanatory variables to account for motion-related artefacts. The output of this GLM was used to estimate the representational similarity between each trial, using the representational similarity analysis toolbox (RSA) (Kriegeskorte et al., 2008a; Nili et al., 2014). The dissimilarity between the response pattern elicited on each trial was estimated using the Mahalanobis distance (Walther et al., 2016), and expressed using correlation distances (1-r). For each participant, the dissimilarity value for the response patterns associated with each trial were represented in each cell of a representational dissimilarity matrix (RDM). Thus, for each stimulus, all trials containing the stimulus were included to estimate a stimulus representation, e.g. trials contributing to the representation of stimulus 1 in memory 1 included all pairs of stimuli shown on a yellow background that included stimulus 1, i.e. 1-1, 1-2, … 1-7. To estimate summary statistics, the Kendall rank correlation coefficient was estimated between the participant's RDM and a model RDM (Supplementary Figure 5.2D-F). These summary statistics were then tested at the group level using a two-sided Wilcoxon signed-rank test across participants. This indicated whether the difference in correlation coefficients between two conditions was greater than zero. This approach allowed for significant within stimulus exemplar discrimination (Supplementary Figure 5.2H-I). To estimate a confusion matrix across memory 1 and 2 (Figure 5.3A), the RDM for each participant was sorted by stimulus type and average representational dissimilarity measure within and between each stimulus was calculated to generate a 7x7 matrix.

5.3.11 Correlations between fMRI, behavioural and MRS data

To assess the relationship between hippocampal BOLD signal and behaviour, a Pearson's correlation was used. Due to outlier data points in the fMRI cross-stimulus suppression measure (see Figure 5.5B-D, Figure 5.6A-B, Supplementary Figure 5.4H), Spearman's rank correlation was used to assess the relationship between fMRI cross-stimulus suppression and changes in GABA or behaviour. Correlations were plotted between standardised residuals, using a partial correlation method to account for unwanted variance attributed to other variables, such as participants' performance on the training task. The partial correlation involved using ordinary-least squares multiple regression to estimate the residuals:

$$Y_i = b_0 + b_1 x_i + \varepsilon_i \quad \text{and} \quad Y_j = c_o + c_1 x_j + \varepsilon_j$$

where Y_i and Y_j represent the variables for which there is hypothesised to be a predictive relationship (e.g. neural and behavioural measures), x_i and x_j represent 'nuisance' variables for which variance is to be accounted (e.g. learning accuracy shown in Supplementary Figure 5.1C-D), b_0 and c_0 represent the intercepts, b_1 and c_1 represent the regression coefficients on the 'nuisance' variables. The partial correlation coefficient was then estimated as the correlation between the resulting residuals, with the mean of the original variables added to the standardised residuals to aid interpretability: $(\varepsilon_i + b_0 + \text{mean}(Y_i))$ and $(\varepsilon_j + c_0 + \text{mean}(Y_j))$.

A permutation test was used to quantify the difference in correlation between behavioural performance and hippocampal BOLD in Block 1 versus Block 2. To estimate a null distribution subject labels for hippocampal BOLD signal were permuted 10,000 times, before being correlated with behavioural performance. The difference in correlation between the Block 1 and Block 2 correlations was then computed for all 10,000 examples. The true difference between Block 1 and Block 2 correlations was compared against the null distribution to generate a p-value (Supplementary Figure 5.6D).

5.4 Results

5.4.1 Associative learning and experimental design

On day 1 of the experiment, participants learned a set of associations between seven rotationally-invariant abstract stimuli (Figure 5.1A) which together formed 'memory 1'. Within memory 1 each stimulus was associated with two other stimuli, giving seven bidirectional associations in total. The set of associations could be arranged into a ring structure (Figure 5.1B), although this was never made explicit to the participants. Instead, participants were instructed to

learn the associations using a three-alternative forced choice task (Figure 5.1D, Supplementary Figure 5.1A,C, see Methods). Rotationally-invariant abstract stimuli were used so that we could later make precise predictions about the brain region sensitive to the learned associations (Barron et al., 2016a).

On day 2 of the experiment, participants learned a second set of bidirectional associations between the same seven abstract stimuli (Figure 5.1A) which together formed 'memory 2'. As in memory 1, each stimulus in memory 2 was associated with two other stimuli (Figure 5.1C). Participants again learned these associations using the three-alternative forced choice task (Figure 5.1E, Supplementary Figure 5.1B,D). Critically, the relational position of the seven stimuli within the ring differed between memory 1 and memory 2, as the position of stimuli 3 and 6 was switched (Figure 5.1B-C). Consequently, four of the seven associations in memory 2 were different from those in memory 1, while three associations remained the same. To help participants distinguish between memory 1 and memory 2, contextual cues were used, consisting of a unique background colour (yellow or blue, randomised across participants) (Figure 5.1D-E) and a time interval of approximately 24 hours between learning sessions (Figure 5.1F).

Thus, memory 1 and memory 2 included the same stimuli but had different relational structures. The difference in relational structure was designed to ensure that a subset of associations across memory 1 and 2, those containing stimuli 3 or 6, were different, while the remaining associations were matched. We predicted that associations containing elements 3 or 6 were susceptible to memory interference, where memory interference manifests as recall of a relational neighbour from the alternative, inappropriate memory. Meanwhile the matched portion of the two memories provided the necessary control. The experimental design therefore included precise and controlled markers of memory interference that could be assessed at both a behavioural and neural level.

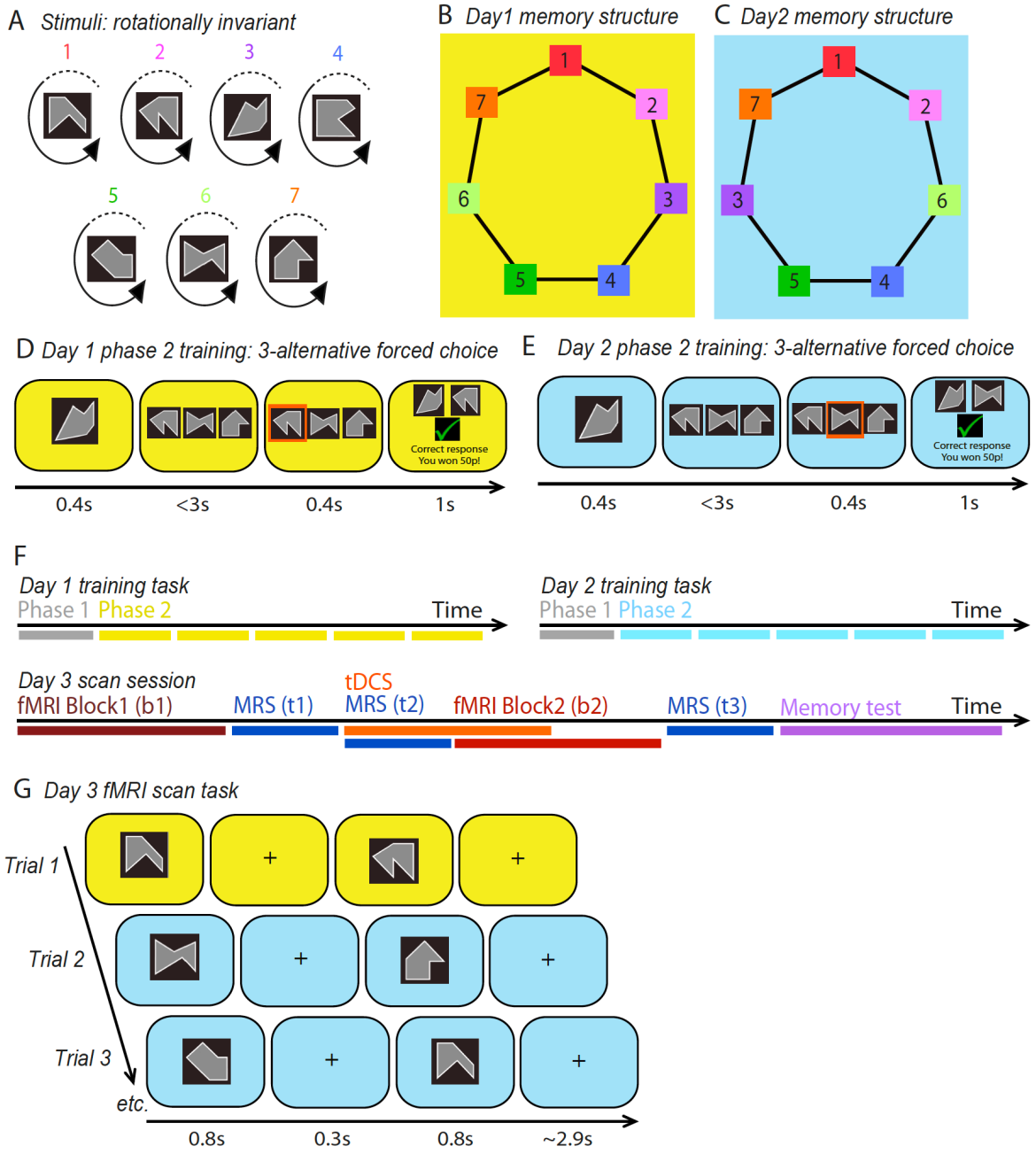


Figure 5.1 Experimental design

A Seven stimuli were used for the experiment, where each stimulus was an abstract shape which could appear in four possible rotations. **B-E** Example experimental protocol: **B** On day 1, participants learned seven associations between pairs of the seven stimuli shown in A, and a contextual cue was provided using a yellow background. The associations could be arranged in a ring structure, although this was not explicitly shown to participants. **C** On day 2, participants learned seven associations between pairs of the seven stimuli, four of which were different from the associations learned on day 1 as the position of stimuli 3 and 6 were swapped. A contextual cue was provided using a blue background, a different colour from that used on day 1. **D-E** To learn the associations on day 1 (D) and day 2 (E), participants performed a 3-alternative forced choice task where the appropriate background colour was used to provide a contextual cue. **F** Schematic showing protocol used across all 3 days of the experiment. On day 1 and 2, participants completed phase 1 (Supplementary Figure 5.1A-B) and at least 5 blocks of phase 2 (Figure 5.1D-E) of the training task. On day 3 of the experiment, participants started with Block 1 ('b1') of the scan task, as shown in Figure 5.1G, before MRS measurements ('t1') were taken to estimate baseline measures of 19 different metabolites. Anodal tDCS was then applied for a total of 20 minutes to induce E/I imbalance, with MRS measurements ('t2') taken during the first 10 minutes before Block 2 ('b2') of the scan task commenced. After Block 2 of the scan task, a third set of MRS measurements ('t3') were obtained before participants exited the scanner to perform a surprise memory test. 'b' indicates block of fMRI acquisition. 't' indicates timepoint of MRS

acquisition. **G** During the scan task, participants observed pairs of stimuli presented consecutively against either a yellow or blue background. All possible pairs of stimuli were presented in a random order.

5.4.2 Hippocampus mediates memory interference using context-dependent relational codes

To identify the physiological mechanisms that protect memories from interference, we first considered the contribution made by the hippocampus. First we sought to show evidence for pattern separation between memory 1 and 2 in the hippocampus, thus building on prior evidence (Bonnici et al., 2012; Huffman and Stark, 2014; Yassa and Stark, 2011). Second, we assessed the organisation of contextual representations that reflect the output of pattern separation. In accordance with the idea that the hippocampus organises representations according to a relational code (Eichenbaum, 2004; Nadel, 2008), we hypothesised that elements that have different relational positions across two overlapping but context-dependent memories are maximally separated. For the paradigm implemented here, we therefore predicted that representations of stimuli 3 and 6 would show maximum pattern separation across memory 1 and 2.

To test these predictions, on day 3 of the experiment we used fMRI to measure the BOLD response to the associative memories learned in memory 1 and 2 (Figure 5.1F). On each trial of the scan task, a pair of stimuli was presented on either a yellow or a blue background to provide a contextual cue for memory 1 or memory 2 (Figure 5.1G, see *Methods*). We controlled for potential confounds introduced by expectation suppression (Summerfield et al., 2008) by ensuring that each possible pair of stimuli was presented equally often in a fully randomised order. To ensure participants paid close attention to the stimuli presented during the scan, participants were instructed to detect 'odd-ball' stimuli which were not part of the seven experienced during training. To improve the signal to noise ratio (SNR) of the fMRI data in brain regions for which we had strong prior hypotheses, we restricted the fMRI sequence to a partial volume, thus increasing the number of measurements acquired (for example partial volume see Supplementary Figure 5.2A-C).

Across two fMRI task blocks ('Block 1' and 'Block 2'), we observed an increase in the hippocampal BOLD signal to pairs of stimuli that had a different relational position across memory 1 and 2 (i.e. pairs of stimuli that included stimuli 3 and 6), relative to pairs of stimuli that had the same relational position across both memories (Figure 5.2A-C: contrast of interest; right hippocampus: t_{23}=4.34, p=0.015, peak-level Family-Wise Error (FWE) corrected using small volume correction (SVC), Figure 5.2A; left hippocampus: t_{23}=3.66, p=0.056, peak-level FWE

corrected using SVC, Figure 5.2B). Therefore, the hippocampal BOLD signal increased on trials where there was opportunity for interference between the two memories.

We next asked whether this hippocampal BOLD signal could predict behavioural measures of memory interference. After the scan session, participants performed a surprise memory test and we assessed recall accuracy for all seven associations within memory 1 and memory 2. The memory test involved the three-alternative force choice task used during training, but now in the absence of feedback (Figure 5.2D). On average, participants correctly recalled the appropriate association on 72.9% of trials, showing higher accuracy for more recent memories (paired-sample t-test, $t25=3.99$, $p<0.001$, Figure 5.2E). In addition to participants' overall memory accuracy, behavioural memory interference was quantified using participants' performance on 'foil trials', namely those trials where the choice stimuli included the stimulus that was correct given the contextual background cue, but also a 'foil' stimulus that would be correct in the alternative memory (Figure 5.2F). The percentage of 'foil errors' made by a participant corresponded to the percentage of foil trials where the foil stimulus was chosen rather than the correct stimulus. Thus, the percentage of foil errors characterised the extent to which participants recalled associative memories learned in context 2 when in context 1 (or vice versa), providing a behavioural measure of memory interference that reflects inappropriate overgeneralisation across contexts. Across participants, we observed a positive relationship between the hippocampal BOLD signal and the percentage of foil errors (hippocampal BOLD signal to pairs of stimuli that included stimuli 3 or 6 minus all other pairs in Block 1, from ROI shown in Figure 5.2G, versus percentage of foil errors: $r_{23}=0.54$, $p=0.006$, Figure 5.2H, after accounting for differences in learning, see Methods).

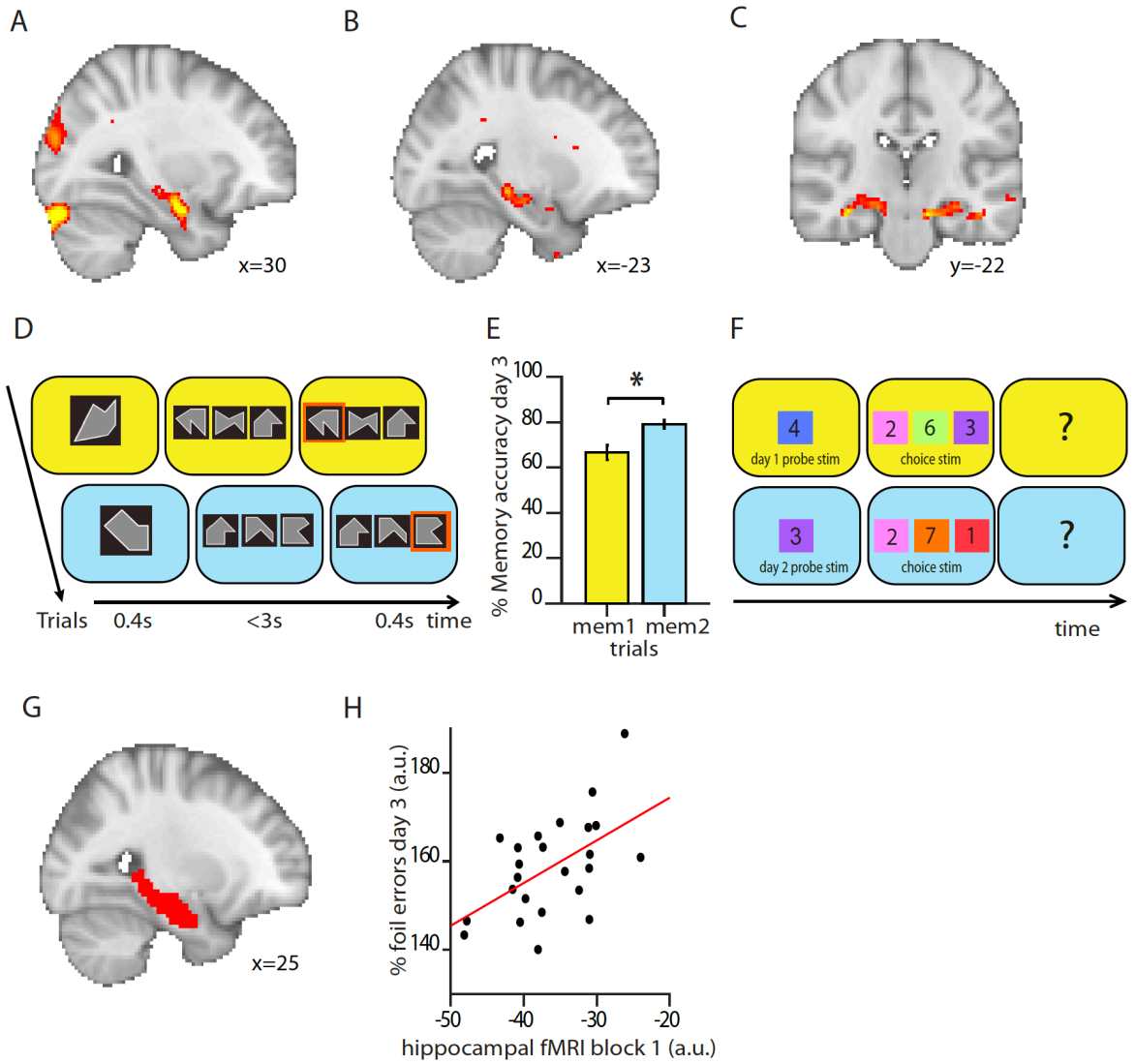


Figure 5.2 The hippocampus mediates memory interference
A-C Hippocampal BOLD signal was higher on trials where there was opportunity for memory interference, i.e. when trials include stimuli 3 or 6 that have a different relational position between memory 1 and 2 (contrast estimated across Block 1 and 2 of the fMRI task, thresholded at $p<0.01$ for visualisation). **A-B** Hippocampal BOLD signal was significantly higher in right hippocampus ($p=0.015$, FWE corrected, peak $t_{23}=4.34$, A), while a similar trend was observed in the left hippocampus ($p=0.056$, FWE corrected, peak $t_{23}=3.66$, B). **C** Hippocampal BOLD signal across both hippocampi, for visualisation. **D** After exiting the scanner on day 3, participants performed a surprise memory test. On each trial participants were presented with a probe stimulus with the background colour providing the contextual cue. They were then presented with three option stimuli and were required to choose the stimulus correctly paired with the probe stimulus in the absence of feedback. **E** Memory accuracy on the surprise memory test (mean, ± SEM), for memory 1 ('mem1') and memory 2 ('mem2'). The more recent associations in memory 2 were remembered more accurately than those in memory 1 (paired t-test: $t_{25}=3.99$, $p<0.001$). **F** Foil trials on the surprise memory test shown in D were identified as those trials where one of the three stimuli was incorrect given the current context, but correct in the alternative context. **G** Hippocampal ROI across both right and left hippocampi, used to perform SVC method in A-B, to extract parameter estimates in H, and to extract activity patterns for RSA in Figure 5.3**. H** The hippocampal BOLD response to trials where there was opportunity for memory interference in Block 1 (Figure 5.1F) predicted behavioural memory interference on the post-scan surprise memory test (Pearson's correlation: $r_{23}=0.54$, $p=0.006$, after accounting for differences in learning, see Methods). 'a.u.' refers to 'arbitrary units'.

While these results suggest that the hippocampus may play a key role in mediating memory interference, they leave open the nature of the hippocampal code. If the hippocampus uses

contextual representations to pattern separate overlapping memories, then representations of stimuli within memory 1 should be more similar to representations of other stimuli within memory 1 compared to representations of stimuli in memory 2. To test this hypothesis we extracted the pattern of activity across voxels in a hippocampal ROI (Figure 5.2G) for each trial in both Block 1 and 2. We then used representational similarity analysis (RSA) to quantify the representational similarity between memory 1 and 2 for each of the seven stimuli, using the Mahalanobis distance (Figure 5.3A, see Methods). We observed higher representational similarity within- versus between-memory (Figure 5.3B; Wilcoxon sign rank test across the group: $Z_{23}=3.34$, $p<0.001$; Model RDM: Supplementary Figure 5.2D). Consistent with previous literature (Bonnici et al., 2012; Huffman and Stark, 2014; Yassa and Stark, 2011), this suggests that stimulus representations in the hippocampus were pattern separated according to contextual information.

Having shown evidence for pattern separation between memory 1 and 2, we next asked how contextual representations for memory 1 and 2 are organised. We predicted that representations of stimuli within memory 1 and 2 are pattern separated in a manner that reflects the overlap in their relational positions. Given the structure of the learned information, the relational position of each stimulus was defined by its neighbours within the ring structures (Figure 5.1B-C). Therefore, only stimuli 3 and 6 had different relational positions across memory 1 and 2. Across memory 1 and 2, representational dissimilarity was higher for stimuli 3 and 6 relative to all other stimuli (stimuli 1,2,4,5, and 7) (Figure 5.3C; Wilcoxon sign rank test across the group: $Z_{23}=2.26$, $p=0.024$; Model RDM: Supplementary Figure 5.2E). Notably, for this cross-memory comparison, only the background colour of the stimuli changed. This result suggests that contextual representations are organised according to the relational overlap of competing memories, where memory elements that have a different relational position across two memories are represented using more distinct neural codes.

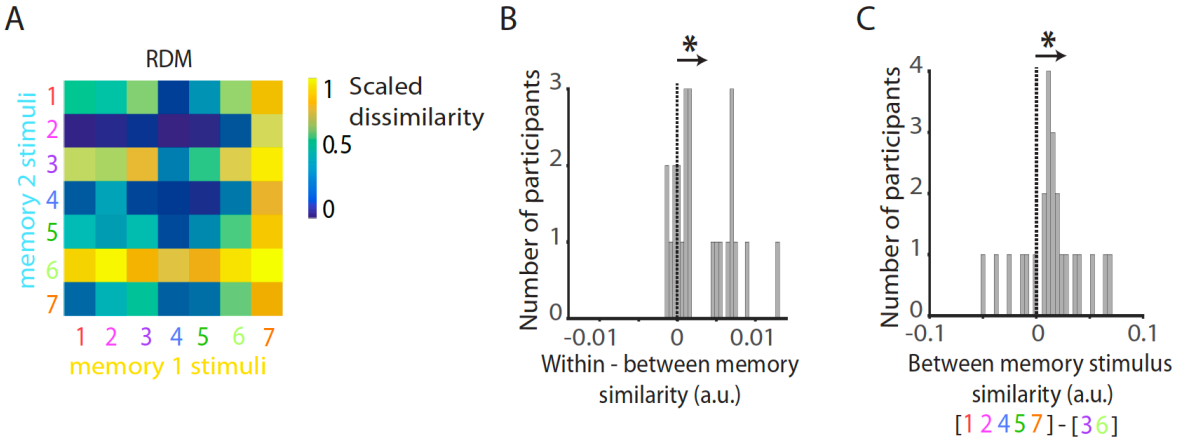


Figure 5.3 Hippocampal representations show evidence for contextual pattern separation, organised by relational overlap

For each trial in the scan task, the pattern of BOLD activity across voxels was extracted from the hippocampus, and representational dissimilarity between all trials containing each of the 7 stimuli was assessed using Representational Similarity Analysis (RSA). For example, representations of stimulus 1 in memory 1 included all pairs of stimuli shown on a yellow background that included stimulus 1, i.e. 1-1, 1-2, … 1-7. **A** Representational dissimilarity matrix (RDM) showing the average representational dissimilarity between stimuli in memory 1 and 2, averaged across all participants, rank transformed and scaled into [0-1] for visualisation. **B** The RDM for each participant was correlated with a model RDM shown in Supplementary Figure 5.2D to test evidence for pattern separation of hippocampal representations by memory. Across participants, significant representational similarity within memory 1/2 versus between memory 1 and 2 was observed (Wilcoxon sign rank test: Z_{23}=3.34, p=<0.001), even if stimuli 3 and 6 were excluded (see Supplementary Figure 5.2G). Note: the dissimilarity of a trial to itself was excluded from the analysis. 'a.u.' refers to 'arbitrary units'. **C** The RDM for each participant was correlated with a model RDM shown in Supplementary Figure 5.2E to test evidence for increased representational dissimilarity of stimuli that had different relative positions across memory 1 and 2 (stimuli 3 and 6). Across participants, significantly greater representational dissimilarity between memory 1 and 2 was observed for stimuli 3 and 6 compared to all other stimuli (Wilcoxon sign rank test: Z_{23}=2.26, p=0.024). 'a.u.' refers to 'arbitrary units'.

5.4.3 Manipulating neocortical E/I balance to measure the effect of inhibition on memory

Having characterised a role for the hippocampus in mediating memory interference, we next asked whether inhibition in the neocortex also plays a key role. Unlike the hippocampus, which appears to encode associated stimuli using conjunctive representations, associative memories in neocortex appear to be stored by excitatory connections that are later balanced by matched inhibition (Barron et al., 2016a; Froemke et al., 2007; Vallentin et al., 2016; Vogels and Abbott, 2009). Therefore, by day 3 of the experiment we expected neocortical representations of memory 1 and 2 to be stored in balanced E/I ensembles. However, if neocortical inhibition plays a critical role in protecting overlapping memories from unwanted interference then it should be possible to induce interference by reducing inhibitory tone. To test this prediction, in the second half of the day 3 scan session we applied non-invasive anodal transcranial direct current stimulation (tDCS) (Figure 5.4A), a tool previously used to induce a transient reduction in the concentration of neocortical GABA (Barron et al., 2016a; Kim et al., 2014; Stagg et al., 2009) and to unmask otherwise silent neocortical associative memories (Figure 5.4B) (Barron et al., 2016a).

Direct current stimulation increases cortical excitability, such that neuronal firing rates increase (Bindman et al., 1962) along with remote motor evoked potentials measured using transcranial magnetic stimulation (TMS) (Nitsche et al., 2005). After stimulation, the increase in cortical excitability is sustained for minutes to hours (Bindman et al., 1962) via a protein synthesis dependent process (Nitsche and Paulus, 2000), which can be used to enhance learning (Jacobson et al., 2011) and recovery from stroke (Hummel and Cohen, 2006). Critically, the mechanism responsible for this increase in cortical excitability appears to involve a reduction in the concentration of available GABA, as evidenced by in vivo spectroscopic measurements (Barron et al., 2016a; Kim et al., 2014; Stagg and Nitsche, 2011; Stagg et al., 2009).

Taking advantage of this non-invasive tool here, we placed the anodal tDCS electrode over anterior Lateral Occipital Complex (LOC) to induce variance in E/I balance, in the target brain region known to encode learned associations for rotationally invariant shapes (Barron et al., 2016a) (Figure 5.4B). The cathodal electrode was placed over the contralateral supraorbital ridge (Figure 5.4A; Supplementary Figure 5.5A-D). Brain stimulation was applied immediately before participants performed a second run of the scan task (Block 2, Figure 5.1F). Before, during and after brain stimulation we used Magnetic Resonance Spectroscopy (MRS) to rapidly measure the concentration of 19 different neural metabolites (Figure 5.1F), including GABA and glutamate, from a 2 x 2 x 2 cm^3 voxel placed just anterior of the anodal electrode (Figure 5.4C). The concentration of each neural metabolite was assessed relative to the concentration of total Creatine (Cr), a suitable reference metabolite. Consistent with previous literature, application of anodal tDCS was accompanied by a significant reduction in the concentration of GABA relative to baseline (GABA:Cr for 'baseline' > 'tDCS', t_{19}=2.32, p=0.016, Figure 5.4E; Supplementary Figure 5.3E). As a consequence, Block 2 of the scan task was performed in a state of E/I imbalance, where excitation outweighed inhibition. The reduction in GABA was not sustained to the period after the task (GABA:Cr for 'baseline' > 'post-task', t_{19}=0.83, p=0.414, Figure 5.4E). In addition to this change in GABA we also observed a significant increase in the concentration of glutamate, but only after the second task session (glutamate:Cr for 'baseline' < 'post-tDCS', t_{19}=2.60, p=0.018; Figure 5.4F). This change in glutamate may be attributed to participants performing Block 2 of the scan task, and doing so in a state of E/I imbalance. See Supplementary Table 5.1 for list of all measured metabolites.

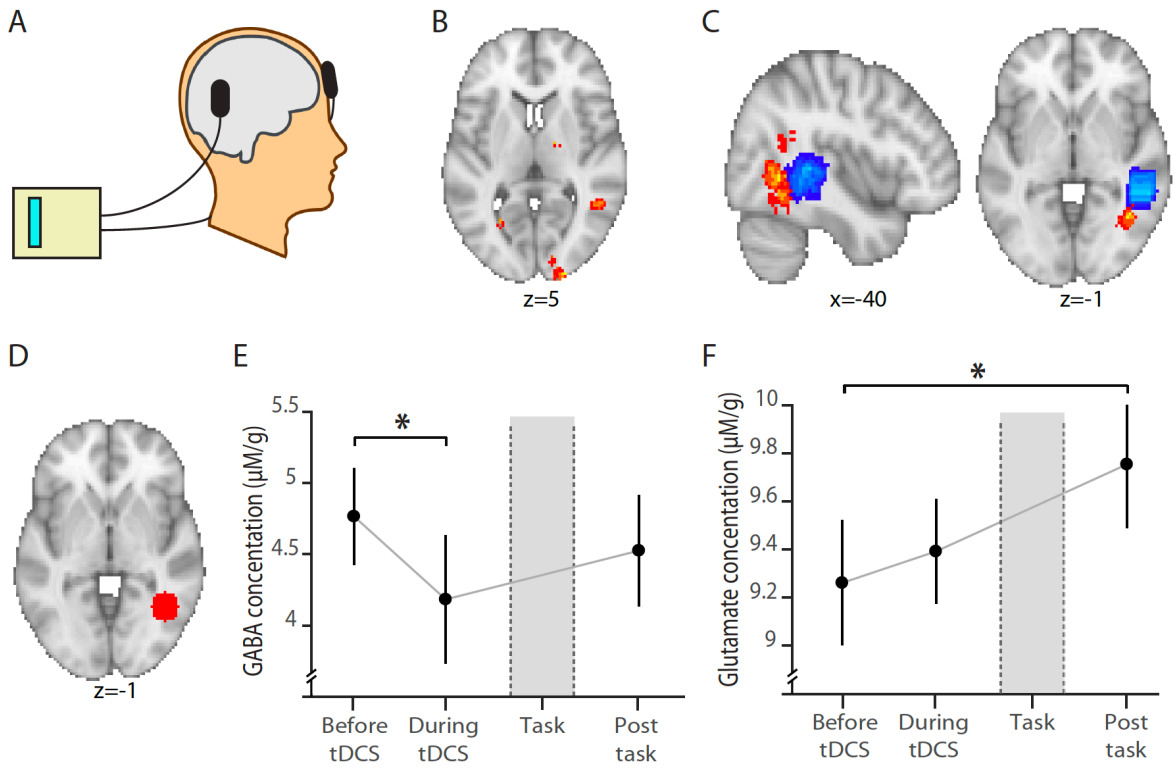


Figure 5.4 Manipulating neocortical E/I balance using brain stimulation unmasks otherwise dormant neocortical memories

A After the first scan task (Block 1), and while participants lay in the scanner, anodal tDCS was applied to the anterior LOC, with the cathodal electrode placed over the contralateral supraorbital ridge. **B** Previously published data (Barron et al., 2016a) showing re-expression of associative memories during application of anodal tDCS. This anterior LOC region was the target location for the anodal tDCS electrode in the current experimental protocol. Orientation: neurological. **C** Blue: Average location of the MRS voxel across participants. Red: Average location of the anodal tDCS electrode across participants, projected into the brain (see *Methods*). Orientation: neurological. **D** 10mm radius sphere defined around the peak tDCS electrode location for all participants (see Methods), used as a region of interest and for small volume correction. Orientation: neurological. **E** MRS was used to quantify the concentration of GABA relative to total Creatine at three time points indicated in Figure 5.1F (shown: mean ±SEM). A significant reduction in relative GABA was observed during tDCS ('Before tDCS' – 'During tDCS', t_{19}= 2.32, p=0.016). **F** MRS was used to quantify the concentration of glutamate relative to total Creatine at three time points indicated in Figure 5.1F (shown: mean ±SEM). A significant increase in relative glutamate was observed after the second scan-task ('Post-task' – 'Before tDCS', t_{19}= 2.60, p=0.018).

5.4.4 Measuring associative memories using cross-stimulus suppression

To test whether anterior LOC inhibition protects memories from interference, we assessed evidence for neural memory interference during the transient period of induced E/I imbalance. To measure memory interference in anterior LOC we sought to index co-activation between representations for different memory elements. We took advantage of fMRI repetition suppression which relies on the fact that neurons show a relative suppression in their activity in response to repeated presentation of a stimulus to which they are sensitive (Miller et al., 1991; Sawamura et al., 2005). While typically used to access sub-voxel representations for single stimuli (Grill-Spector et al., 2006), 'cross-stimulus' suppression can be used to index the relative co-activation or overlap between representations coding for two different stimuli (Barron et al., 2016b). We contrasted the

BOLD response for each pair of stimuli where suppression was expected against the BOLD response to a control pair where suppression was not expected, thus controlling for attentional effects (Figure 5.5A, Supplementary Figure 5.4A). The ring topology of memory 1 and 2 provided an efficient way to ensure that each stimulus contributed to both trials where suppression was expected (directly associated stimuli in one or both contexts) and control trials where suppression was not expected (stimuli separated by up to three associations in both contexts).

We first replicated our previous findings (Barron et al., 2016a), showing that cross-stimulus suppression increases during anodal tDCS between directly associated stimuli that remain the same across memory 1 and 2 (Supplementary Figure 5.4C). Furthermore, this increase in cross-stimulus suppression can be predicted by the relative decrease in GABA concentration (Supplementary Figure 5.4H). This implies that inhibition in anterior LOC acts to quench memory expression of recently acquired associative memories, but during periods of E/I balance these otherwise dormant memories are re-expressed. Interestingly, the extent to which associative memories were re-expressed during E/I imbalance was significant for memory 2 but not memory 1 alone (Supplementary Figure 5.4D-G). This difference between recent and more remote memories may in part be explained by a difference in the strength of associations in memory 1 compared to memory 2, which could be observed at a behavioural level (paired-sample t-test, $t_{25}=3.99$, $p<0.001$, Figure 5.2E), even for associations that remained the same across memory 1 and 2 (paired-sample t-test, $t_{25}=2.16$, $p=0.040$, Supplementary Figure 5.4I).

5.4.5 Memory interference increases during periods of E/I imbalance

Having replicated our previous findings (Barron et al., 2016a), we went on to investigate whether neocortical inhibition plays a critical role in protecting against memory interference. Capitalising on the inter-subject variability to the anodal tDCS manipulation (Figure 5.4E), we predicted a two-way relationship between the drop in relative GABA, our neural measures of memory interference and behavioural measures of memory interference: if neocortical inhibition protects memories from interference, the drop in relative GABA should predict neural measures of memory interference, which should in turn predict behavioural measures of memory interference.

As neural memory interference manifests as activation of a relational neighbour from the alternative, inappropriate memory, we sought to index this unwanted activation using cross-stimulus suppression. To this end, we identified trials during the scan task where participants were shown two stimuli that were unassociated given the memory indicated by the contextual cue, but

directly associated in the alternative memory. During periods of E/I imbalance we predicted an increase in cross-stimulus suppression on these trials, relative to trials where the presented stimuli were indirectly associated in both memory 1 and 2 (Figure 5.5A). Thus, an increase in this cross-stimulus suppression measure provided a proxy for an increase in neural memory interference. Using this measure of memory interference we assessed evidence for the predicted two-way relationship between the drop in concentration of relative GABA, neural memory interference and behavioural memory interference.

Firstly, we considered the relationship between the drop in relative GABA during application of anodal tDCS and the increase in neural memory interference from the first to the second fMRI scan task block. Across participants, the drop in relative GABA positively predicted the increase in neural memory interference measured using cross-stimulus suppression across memory 1 and 2 (r_{17}=0.55, p=0.021, Figure 5.5B, after accounting for changes in glutamate, see Methods). Cross-stimulus suppression measured from participants with minimal change in the concentration of relative GABA thus provided effective parametric control for participants where a larger drop in relative GABA was observed, mitigating the need for a separate sham condition. Thus, the variation in the drop in relative GABA observed across participants provided a stringent framework in which to assess the effect of E/I imbalance on cross-stimulus suppression. This result suggests that interference between overlapping memories is predicted by E/I imbalance.

Secondly, we considered the relationship between neural and behavioural memory interference. Taking behavioural performance from the surprise memory test performed after the scan, we predicted a positive relationship between neural memory interference and the percentage of foil errors on the memory test, but a negative relationship between neural memory interference and overall accuracy on the memory test. Consistent with these predictions, our cross-stimulus suppression index for neural memory interference (Block 2 – Block 1) positively predicted the percentage of foil errors, and negatively predicted overall behavioural memory accuracy (fMRI vs. foil errors: r_{17}=0.58, p=0.013, Figure 5.5C; fMRI vs. overall accuracy: r_{17}=-0.67, p=0.003, Figure 5.5D; after accounting for differences in learning and changes in relative GABA and glutamate, see Methods). In summary, participants who showed greater cross-stimulus suppression during periods of E/I imbalance also made more errors.

Together these results suggest that a reduction in neocortical GABAergic tone leads to an increase in neural memory interference which manifests in behaviour as an increase in memory errors. While this two-way relationship capitalises on the variability observed across participants, we next asked whether there was a main effect of anodal tDCS on neural memory interference.

Using cross-stimulus suppression as a proxy for memory interference, we predicted an overall increase in neocortical memory interference during the application of anodal tDCS. Furthermore, given that a reduction in neocortical GABA resulted in pronounced re-expression of more recent associations in memory 2 (Supplementary Figure 5.4F,G), we predicted memory interference would manifest in memory 1 due to expression of associations in memory 2 intruding or overriding the appropriate expression of associations in memory 1.

To maximise sensitivity to the effect of anodal tDCS, we tested for memory interference using cross-stimulus suppression within an ROI defined from the peak anodal tDCS electrode location, averaged across all participants (Figure 5.4C, see Methods). Within this ROI, across both memory 1 and 2 we observed a trend towards an increase in cross-stimulus suppression during application of anodal tDCS (Block 2 > Block 1, t_{23}=1.79, p=0.087, Supplementary Figure 5.6A). However, consistent with our prediction, for memory 1 but not memory 2 there was a pronounced increase in our cross-stimulus suppression measure of memory interference (Block 2 > Block 1, memory 1: t_{23}=3.05, p=0.006 Figure 5.5E-F, Supplementary Figure 5.6B; memory 2: t_{23}=0.57, p=0.573, Supplementary Figure 5.6C). To confirm that memory interference was observed during application of anodal tDCS, we also assessed effects in Block 2 alone. Within the same ROI we again observed significant cross-stimulus suppression for memory 1 but not memory 2 (memory 1: t23=3.00, p=0.006, Figure 5.5E-G, Supplementary Figure 5.6B), and within a 10 mm sphere centered on the peak of the average anodal tDCS electrode location (memory 1: t_{23}=3.60, p=0.027, peak-level FWE corrected using SVC with ROI shown in Figure 5.4D). Critically, this cross-stimulus measure of interference in memory 1 was anatomically proximal to the re-expression of directly associated memories in memory 2 reported previously (Figure 5.5H).

These results suggest that re-expression of directly associated stimuli in memory 2 leads to interference with overlapping, but contextually distinct, associations in memory 1. In a final analysis, we asked whether the differential strength of memory 1 and 2 at encoding also predicts neural memory interference during periods of E/I imbalance. We found that participants' average learning accuracy for associations in memory 2 positively predicted the cross-stimulus suppression measure for memory interference (r_{17}=0.68, p=0.003, Figure 5.6A, after accounting for differences in learning on day 1, memory accuracy and changes in relative GABA and glutamate, see Methods), while a trend towards the reverse relationship was observed for memory 1 (r_{17}=-0.47, p=0.053, Figure 5.6B, after accounting for differences in learning on day 2, memory accuracy and changes in relative GABA and glutamate, see Methods). Therefore, participants that weakly encoded memory 1 but strongly encoded memory 2 were more prone to memory interference. Together with results above, this suggests that interference between two memories during periods of E/I

imbalance can be predicted by the extent to which E/I imbalance is induced (Figure 5.5B), but also the relative strength of the memories at encoding (Figure 5.6A-B).

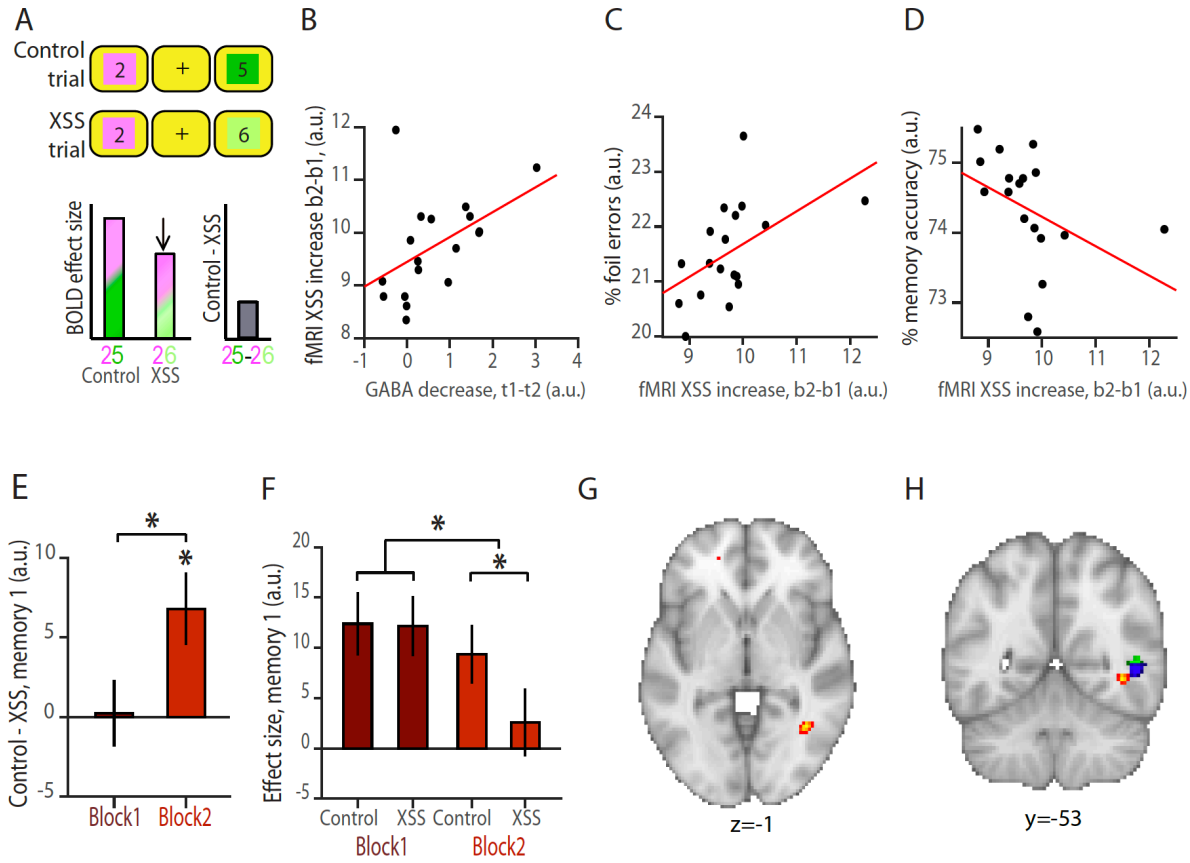


Figure 5.5 Memory interference for more remote memories increases with brain stimulation
In all panels, 'a.u.' refers to 'arbitrary units'; XSS indicates cross-stimulus suppression; 'b' indicates block of fMRI acquisition, as shown in Figure 5.1F; 't' indicates 'timepoint' of MRS measurement acquisition, as shown in Figure 5.1F. **A** When participants performed the scan task in E/I imbalance, we predicted an increase in cross-stimulus suppression ('XSS') on trials where participants observed pairs of stimuli that were unassociated in the current context but directly associated in the alternative context, relative to control trials where participants observed pairs of stimuli that were unassociated in both contexts. This difference between 'Control' and 'XSS' trials could be measured using the BOLD signal and was used as an index for neural memory interference. **B** Across participants, the decrease in relative GABA concentration observed during tDCS ('Before tDCS' – 'During tDCS', Figure 5.4E) positively predicted the increase in fMRI cross-stimulus suppression measure for interference between memory 1 and 2 ('Control' – 'XSS' for Block 2 – Block 1, Figure 5.5A) (Spearman correlation: r_{17}=0.55, p=0.021, after accounting for changes in glutamate, see Methods). **C** Across participants, the increase in cross-stimulus suppression used to measure memory interference ('Control' – 'XSS' for Block 2 – Block 1, Figure 5.5A) positively predicted the percentage of foil errors participants made on the surprise memory test on day 3 (Spearman correlation: r_{17}=0.58, p=0.013, after accounting for differences in learning and changes in relative GABA and glutamate). **D** Across participants, the increase in cross-stimulus suppression used to measure memory interference ('Control' – 'XSS' for Block 2 – Block 1) negatively predicted average memory accuracy on the surprise memory test on day 3 (Spearman correlation: r_{17}=-0.67, p=0.003, after accounting for differences in learning and changes in relative GABA and glutamate). **E** Within an ROI defined from the peak average tDCS electrode location shown in Figure 5.4C, extracted parameter estimates for memory 1 (shown: mean ±SEM) revealed a significant increase in the fMRI cross-stimulus suppression measure for memory interference ('Control' – 'XSS', as shown in A) from Block 1 to 2 and during Block 2 alone ('Control' – 'XSS' for Block 2 – Block 1: t_{23}=3.05, p=0.006; 'Control' – 'XSS' for Block 2: t_{23}=3.00, p=0.006). **F** Extracted parameter estimates from E split into the 'Control' and 'XSS' conditions, as described in *A* (shown: mean ±SEM). **G** T-statistic map for cross-stimulus suppression index for neural memory interference during Block 2 between unassociated stimuli in memory 1 that are directly associated in memory 2, relative to pairs of stimuli that are unassociated in both memories, thresholded at p<0.01 uncorrected for visualisation. Orientation: neurological. **H** Illustrating the anatomical proximity between the effects shown in

Figure 5.5G (red), Supplementary Figure 5.4G (blue) and previously acquired data set shown in Figure 5.4B (green). Orientation: neurological.

5.4.6 The interplay between the hippocampus and anterior LOC

These data suggest that in addition to hippocampal pattern separation mechanisms, neocortical inhibition plays a key role in protecting memories from interference. To assess the interplay between the hippocampal and neocortical mechanisms, we reconsidered the relationship between our neural and behavioural measures of memory interference. We noted that behavioural performance on the surprise memory test after the scan session was predicted by both the hippocampal BOLD signal prior to anodal tDCS (Figure 5.2H), and the change in neocortical cross-stimulus suppression observed during anodal tDCS (Figure 5.5C-D). We asked whether hippocampal BOLD during anodal tDCS (block 2) also predicted behavioural performance. Unlike hippocampal BOLD prior to anodal tDCS (block 1), we observed no relationship between hippocampal BOLD during tDCS (block 2) and behavioural performance (hippocampal BOLD Block 2 vs. foil errors: r_{23}=-0.06, p=0.764, Figure 5.6C, after accounting for differences in learning). Furthermore, this correlation between Block 2 hippocampal BOLD and behaviour was significantly different from the correlation observed between Block 1 hippocampal BOLD and behaviour (difference in correlation coefficient, Block 1 vs Block 2, permutation test: p=0.032, Supplementary Figure 5.6D, see *Methods*).

While this difference in the relationship between behavior and hippocampal BOLD signal in Block 1 versus Block 2 may be explained by the effect of time, this seems unlikely as there was no significant change in the magnitude of the hippocampal BOLD signal from Block 1 to Block 2 (paired t-test: t_{23}=0.25, p=0.802). Rather, these results suggest that in the absence of brain stimulation, the degree to which irrelevant associative memories are represented in the hippocampus predicts memory interference. But, when neocortical GABAergic tone is reduced, signatures of neural memory interference in anterior LOC but not hippocampus predict memory interference. The hippocampus and anterior LOC thus appear to employ distinct mechanisms to mediate memory interference.

5.4.7 Inducing behavioural memory interference using brain stimulation

In a final set of experiments we asked whether application of anodal tDCS alone might be sufficient to induce behavioural measures of memory interference. To this end we repeated the experiment in three additional groups of participants: (2) anodal tDCS or (3) sham-anodal tDCS

(delivered using a double-blind set-up, see Methods), or (4) no intervention. These three additional groups of participants performed the same set of tasks as participants receiving MRI (group 1), but outside the scanner. On the day 3 surprise memory test we observed a significant difference between groups in mean accuracy and in the percentage of normalised foil errors using a one-way ANOVA (mean accuracy: $F_{82}=6.54$, $p<0.001$, Supplementary Figure 5.1E; normalised foil errors: $F_{82}=4.39$, $p=0.007$, Figure 5.6D), and a significant effect of stimulation when using multiple regression to control for variation in learning accuracy and gender that occurred by chance across the four experimental groups (effect of stimulation on mean accuracy: $t_{81}=2.96$, $p=0.004$; normalised foil errors: $t_{81}=2.87$, $p=0.005$). Post-hoc t-tests revealed a significant difference between participants who received both anodal tDCS and MRI compared to participants who did not receive any intervention (group 1, 'MRI', vs. group 4, 'Behav', foil errors: $t_{44}=2.89$, $p=0.006$; group 2, 'tDCS', vs. group 4, 'Behav', foil errors: $t38=2.30$, $p=0.027$; Figure 5.6D). For participants who received sham-stimulation, there was no significant difference in behavioural performance compared to participants who did not receive any intervention (group 3, 'Sham', vs. group 4, 'Behav', foil errors: $t_{38}=1.07$, $p=0.292$; Figure 5.6D). While these results suggest that anodal tDCS increased memory interference at the behavioural level, there was notably no significant difference in behavioural performance between participants who received anodal tDCS and those who received sham stimulation (group 2, 'tDCS', vs. group 3, 'Sham': foil errors: $t_{38}=0.66$, $p=0.515$; Figure 5.6D). Similar results were obtained when using multiple regression to assess differences in the percentage of foil errors while controlling for variation in learning accuracy and gender that occurred by chance across the experimental groups ('MRI' vs 'Behav': $t_{41}=2.93$, $p=0.005$; 'tDCS' vs 'Behav': $t_{35}=2.32$, $p=0.026$; 'Sham' vs. 'Behav': $t_{35}=1.38$, $p=0.177$; 'tDCS' vs 'Sham': $t_{35}=0.59$, $p=0.559$), and when assessing differences in overall memory accuracy (Supplementary Figure 5.1E). These results suggest that while anodal tDCS can induce behavioural memory interference, the expectation of anodal tDCS has a similar effect on some participants. Therefore, rather than mere application of brain stimulation, the change in the concentration of relative GABA and neural measures of memory interference are necessary to reliably predict behavioural measures of memory interference.

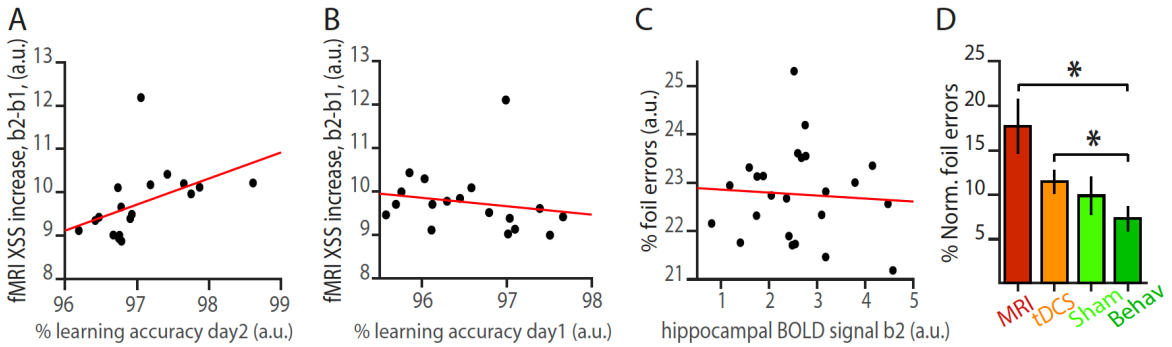


Figure 5.6 Relating learning accuracy to neural measures of memory interference
In all panels, 'a.u.' refers to 'arbitrary units'; XSS indicates cross-stimulus suppression; 'b' indicates block of fMRI acquisition, as shown in Figure 5.1F. **A** Across participants, learning accuracy for memory 2 (on day 2) positively predicted the cross-stimulus suppression index for neural memory interference (Spearman correlation: r_{17}=0.68, p=0.003). **B** Across participants, learning accuracy for memory 1 (on day 1) showed a negative trend with the cross-stimulus suppression index for neural memory interference (Spearman correlation: r_{17}=-0.47, p=0.053). **C** During application of anodal tDCS in the second scan task (Block 2, 'b2'), the hippocampal BOLD response to trials where there was opportunity for memory interference did not predict behavioural measures of memory interference (Pearson correlation: r_{23}=-0.06, p=0.764). **D** Both groups of participants (see Methods) who received tDCS ('MRI' and 'tDCS') showed higher normalised foil errors on the surprise memory test on day 3, relative to participants who received no intervention ('Behav') (two-sample t-test: t_{44}=2.89, p=0.006) (shown: mean ±SEM). However, there was no difference in the percentage of foil errors made by participants who received 'tDCS' and 'Sham' (two-sample t-test: t_{38}=0.66, p=0.515). Normalised foil errors were defined as the percentage of foil errors on foil trials, after subtracting the percentage of non-foil errors on foil trials.

5.5 Discussion

Our past experiences often overlap in their content, but can nevertheless be selectively recalled without interference from other memories. Here, we investigated the neural mechanisms that help protect memories from interference. By training human participants to encode two context-dependent overlapping memories, memory 1 and 2, we reveal evidence for two distinct neural mechanisms that help mitigate memory interference. The first mechanism involves the hippocampus, where overlapping but context-dependent memories are pattern separated according to their relational overlap. The second mechanism involves neocortical inhibition, which protects against unwanted co-activation between neocortical representations. We discuss these two mechanisms in turn, before considering how they may interact.

In the hippocampus we observed an increase in hippocampal BOLD signal when participants observed pairs of stimuli that had different relative positions across memory 1 and 2. In the absence of brain stimulation this BOLD signal predicted participants' performance on a surprise memory test completed after the scan. When we investigated the nature of the underlying hippocampal representations, we found evidence for pattern separation by context, where representations of stimuli were more similar to representations of other stimuli within the same memory compared to representations of stimuli in the alternative memory. This finding is in

agreement with a large body of evidence suggesting an important role for the hippocampus in pattern separation (Yassa and Stark, 2011), a mechanism that is likely mediated by orthogonal contextual representations (Butterly et al., 2012; McKenzie et al., 2014).

But in addition, we show that pattern separation appears to be organised according to the relational overlap between memory 1 and 2, as representations of stimuli 3 and 6 were more dissimilar between memory 1 and 2, compared to all other stimuli. Therefore, pattern separation is enhanced for elements that change relational position between competing memories. These results suggest that contextual representations, which emerge from a competitive pattern separation mechanism, may be analogous to a series of cognitive maps, where each set of learned relationships is stored in a unique map distorted by the relational overlap with competing maps. Interestingly, this account is consistent with the idea that the hippocampus represents a successor representation where stimuli that predict different future states have more distinct representations (Dayan, 1993; Garvert et al., 2017; Momennejad et al., 2017; Stachenfeld et al., 2017). In this light, contextual representations within the hippocampus may be construed as configural representations (Nadel, 2008), where the same machinery responsible for coding spatial relationships is employed when representing abstract, non-spatial stimuli.

While the hippocampus may help minimise interference by separating context-dependent memories according to their relational overlap, the sensory neocortex appears to employ a different mechanism. By down-regulating the concentration of neocortical GABA using anodal tDCS (Barron et al., 2016a; Kim et al., 2014; Stagg et al., 2009), here we show that during periods of E/I imbalance, neocortical memory interference increases. To quantify neural memory interference we used ultra-high field 7T MRI to measure cross-stimulus suppression, a proxy for representational similarity between different elements of the memories (Barron et al., 2016b; Krekelberg et al., 2006). We show that the drop in GABA quantified using MRS predicts our neural measure of memory interference, which in turn predicts behavioural measures of memory interference. This two-way relationship reveals a key role for neocortical inhibitory engrams in protecting against interference.

We note that in this study we used tDCS as a tool to induce variability in the concentration of GABA in anterior LOC, which our MRS measures show was successfully achieved. Indeed, the range in inter-subject variability in induced E/I imbalance provided a stringent framework within which to test our hypotheses, as fMRI measures from participants with a low change in GABA parametrically controlled for fMRI measures from participants with a higher change in GABA. The variability in GABA thus mitigated the need for a sham control group, or a control voxel from

which to measure MRS, and provided a precise prediction for neural and, in turn, behavioural memory interference. When we assessed the effect of tDCS alone on behavioural memory interference, no significant difference in behavioural measures of memory interference was observed between a group of participants receiving tDCS relative to a group receiving sham stimulation. These behavioural findings confirm that a measure of E/I imbalance, and not mere application of tDCS, is necessary to provide a reliable predictor of memory interference.

While our findings suggest that inhibitory engrams are critical for stable memory storage, they also raise a number of questions regarding the formation of inhibitory engrams and the accompanying timescale of this process. In the rodent primary auditory cortex, changes in the strength of excitatory connections are accompanied by inhibitory rebalancing after approximately 90 minutes (Froemke et al., 2007). This implies that a 'critical period' of E/I imbalance and memory instability occurs between initial learning and the formation of inhibitory engrams. Consistent with this hypothesis, a transient period of memory instability has been reported immediately after learning, during which memories may be integrated with existing knowledge that share abstract or higher-level properties (Mosha and Robertson, 2016). This integration is facilitated by offline reactivation and coordinated interactions between hippocampal and neocortical engrams (Mosha and Robertson, 2016; Preston and Eichenbaum, 2013; Schlichting and Preston, 2014).

While this opportunity to integrate newly encoded memories with existing knowledge has clear advantages, the relative instability of memories during this 'critical period' makes them vulnerable to interference. This trade-off between integration and interference may determine the transient nature of the 'critical period'. Indeed, if sufficient time is left between acquisition of memories that are overlapping or share a common structure, integration is no longer observed (Mosha and Robertson, 2016), nor is interference, as shown here in the absence of anodal tDCS stimulation. In addition to time, other factors such as overlearning also appear to terminate the 'critical period' (Mosha and Robertson, 2016) by restoring E/I balance with a shift from glutamate-dominated excitation to GABA-dominated inhibition (Shibata et al., 2017).

But in addition to the timescale of inhibitory engram formation, both memory expression and memory interference are also likely to be affected by the underlying strength of the encoded memory. For example, if the excitatory strength of an associative memory is weak, neural and behavioural expression of a memory during recall will be reduced. Here, we show a significant reduction in behavioural measures of memory accuracy for associations in memory 1 relative to memory 2 (Figure 5.2E), even for associations that remain the same across memory 1 and 2

(Supplementary Figure 5.4I). At the neural level, a relative weakening in the strength of excitatory connections in memory 1 relative memory 2 may explain why cross-stimulus suppression between directly associated stimuli, an index for memory expression, was only observed for memory 2 (Supplementary Figure 5.4D-E), while memory interference effects were only observed for memory 1 (Figure 5.5E-G).

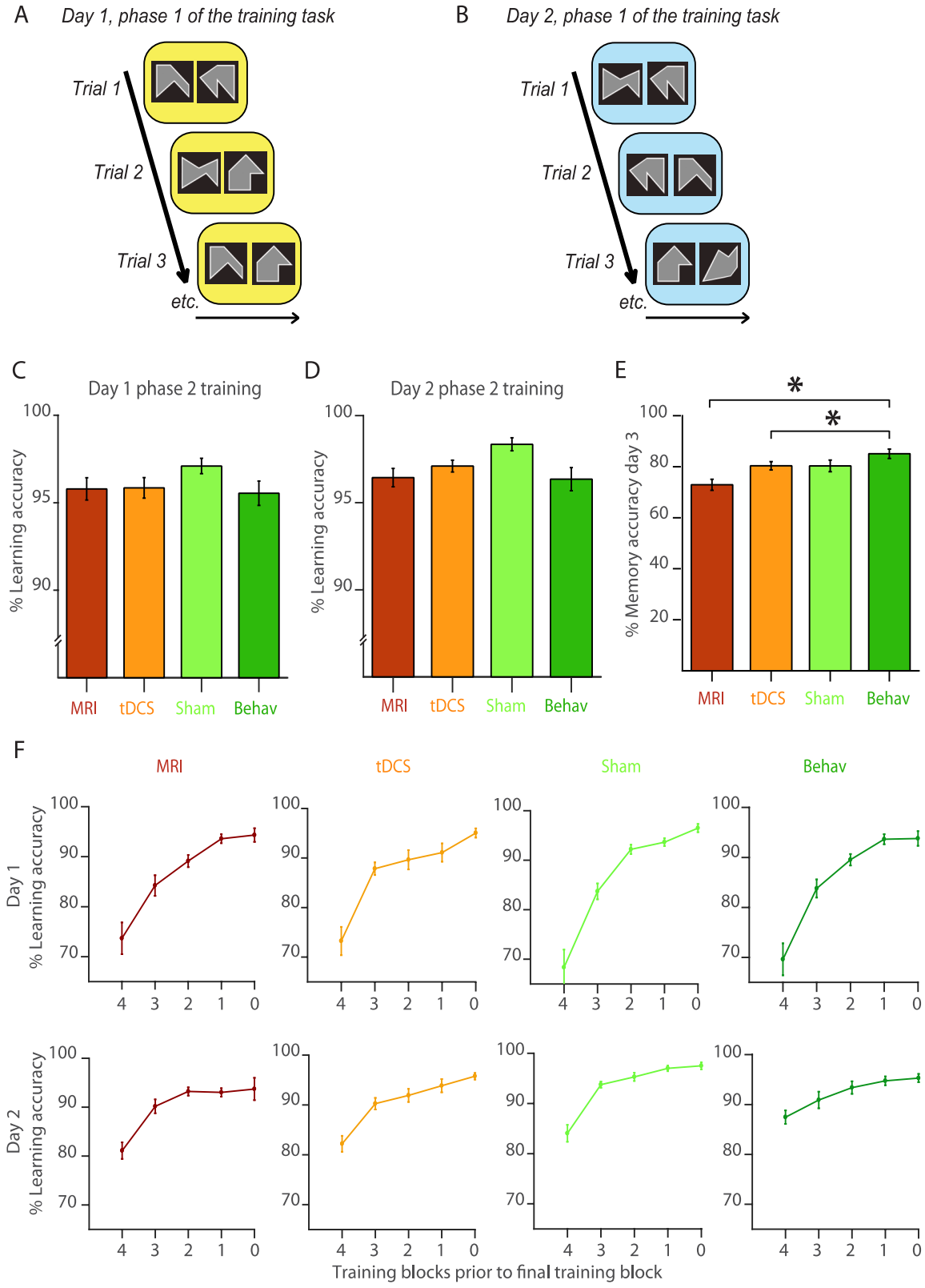
By combining ultra-high field 7T fMRI with MRS, brain stimulation and behavioural measures, the protocol described here illustrates how macroscopic measures of the human brain can be used to index micro-circuit processes. This has notable translational value for clinical populations where microcircuit disruption is not readily amenable to investigation, particularly conditions that have been attributed to disturbances in E/I balance. For example, in schizophrenia delusions and hallucinations have been attributed to perturbed inhibitory gating (Vogels and Abbott, 2007; Yizhar et al., 2011), while memory loss and confusion in early stage dementia have been associated with hyperactivity (Busche and Konnerth, 2016). The data presented here may be considered a model for these clinical phenotypes, where neocortical E/I imbalance causes unwanted reactivation of irrelevant memories that may have overlapping excitatory engrams with those activated by incoming stimuli. In the absence of appropriate inhibitory regulation, otherwise latent memory engrams are activated in an uncontrolled manner causing confusion or hallucinations. The protocol implemented here thus provides a basis from which to further explore mechanisms responsible for memory impairment in clinical conditions that report evidence for E/I imbalance.

Finally, we consider the interplay between the hippocampal and neocortical mechanisms for mitigating against memory interference. When we reduced the concentration of neocortical GABA using anodal tDCS, interestingly the hippocampal BOLD signal no longer predicted behavioural performance on the surprise memory test. This suggests that inhibitory gating in the anterior LOC may influence context-dependent hippocampal representations. Although beyond the scope of this investigation it is also interesting to speculate about the reverse relationship, the influence of context-dependent relational hippocampal codes on neocortical memory engrams. One possibility is that analogous to cholinergic modulation of neocortical interneurons observed during context-dependent behaviour (Kuchibhotla et al., 2017), the hippocampus may mediate selective release of neocortical memory engrams by targeting neocortical inhibition (Barron et al., 2017). This interaction between the hippocampus and neocortex may facilitate selective reactivation of neocortical representations during memory recall, providing an index for distributed memory engrams (Teyler and Rudy, 2007).

5.6 Conclusion

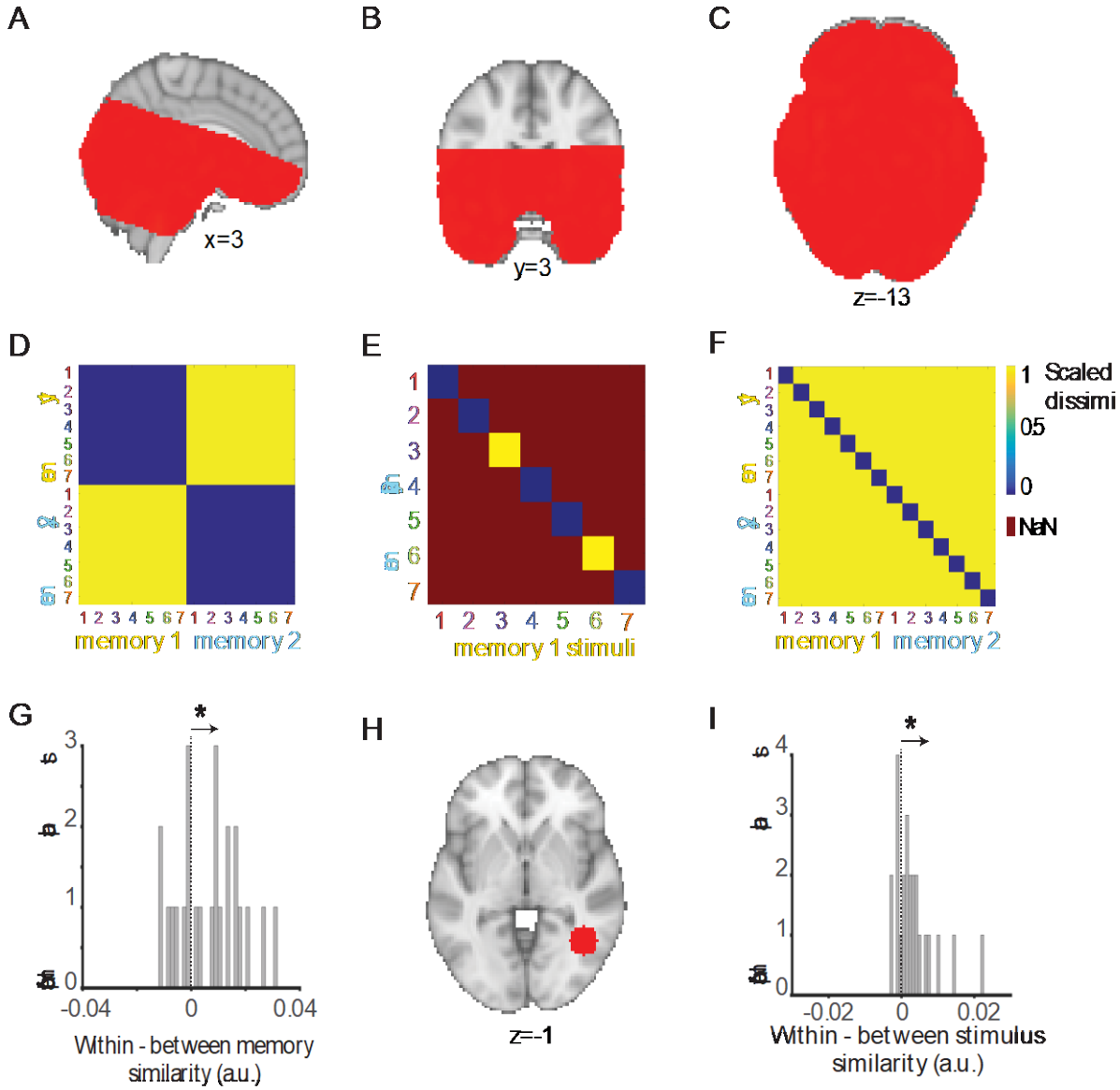
In this chapter I showed that both hippocampal relational codes and cortical inhibition prevent interference, or overgeneralisation, between overlapping memories. The finding that hippocampal activity predicts neocortical memory interference only when no tDCS was applied, suggests that when E/I balance is disturbed, hippocampal-neocortical communication is disrupted, which may underlie maladaptive interference.

5.7 Supplementary figures



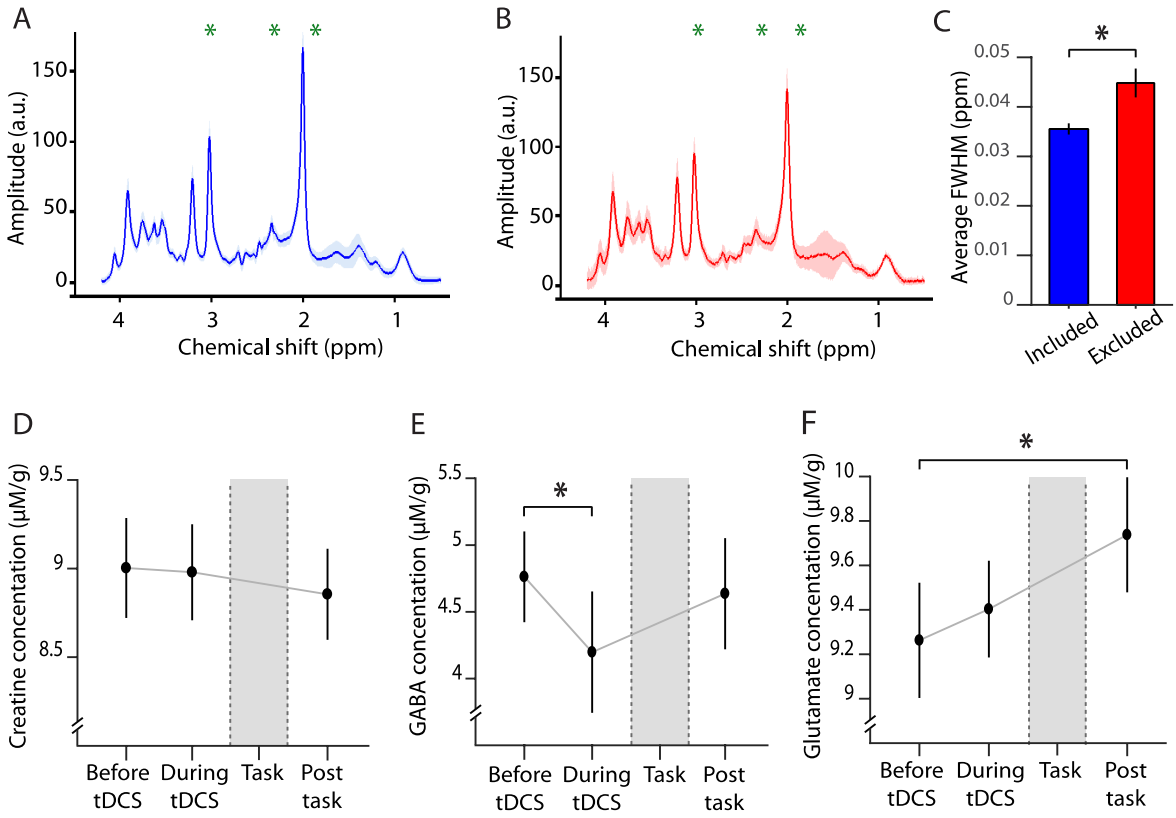
Supplementary Figure 5.1 Training task and surprise memory test
A-B The training task performed by participants on day 1 and 2 of the task involved two phases. Phase 1 involved a passive task (shown here) while phase 2 involved an active task (shown in Figure 5.1D-E). On each trial of phase

1 of the training task participants were shown a pair of associated stimuli for 3 s duration. The stimulus that appeared on the left-hand side of the screen was randomised. Each pair of stimuli was presented 4 times in total and the order in which pairs of associated stimuli were presented was randomised across trials. The background colour on the screen provided a contextual cue for each day of the training task and was either yellow (A) or blue (B). Participants were required to passively observe the pairs of stimuli and instructed to try to learn the associative pairings. **C-D** Percentage learning accuracy during phase 2 of the training task (Figure 5.1D-E) for the four different groups of participants (see Methods) on day 1 (C) and day 2 (D). Shown: mean ±SEM. Some participants got tired during training and showed a reduction in task performance during the final and/or penultimate task block. For this reason, 'learning accuracy' was estimated as the average performance across trials on each participant's highest performing task block. On day 1 there was no significant difference in learning accuracy between any pair of the four experimental groups. However, on day 2, a significant difference in learning accuracy was observed between the 'sham' group and the 'MRI' group (t_{44}=2.12, p=0.040), but not between any other groups. This difference occurred by chance as the allocation of participants to the 'tDCS' and 'sham' group was double blinded. **E** Accuracy on the surprise memory test on day 3 for the four different groups of participants. Shown: mean ± SEM. We observed a significant difference between groups in mean memory accuracy using a one-way ANOVA (mean accuracy: F_{82}=6.54, p<0.001), and a significant effect of stimulation when using multiple regression to control for variation in learning accuracy and gender that occurred by chance across the four experimental groups (effect of stimulation on mean accuracy: t_{81}=2.96, p=0.004). Post-hoc t-tests revealed significantly lower overall memory accuracy for participants who received tDCS and MRI relative to participants who received no intervention ('MRI' vs. 'Behav': t_{44}=4.11, p<0.001), with a similar trend for participants who received tDCS without MRI ('tDCS' vs 'Behav': t_{38}=1.91, p=0.064). However, there was no difference in performance between the 'tDCS' and 'Sham' groups ('tDCS' vs 'Sham': t_{38}=0.02, p=0.986). Similar results were obtained when using multiple regression to control for variation in learning accuracy and gender that occurred by chance between the four experimental groups ('MRI' vs 'Behav': t_{41}=5.17, p<0.001; 'tDCS' vs 'Behav': t35= 2.19, p=0.035; 'tDCS' vs 'Sham': t35=0.568, p=0.574). **F** Percentage learning accuracy during phase 2 of the training task (Figure 5.1D-E) for the four different groups of participants (see *Methods*), split by training block for both day 1 (upper row) and day 2 (lower row). Shown: mean ±SEM. Performance accuracy is shown for the final 5 training blocks as all participants were trained until they completed at least 5 training blocks (see Methods).

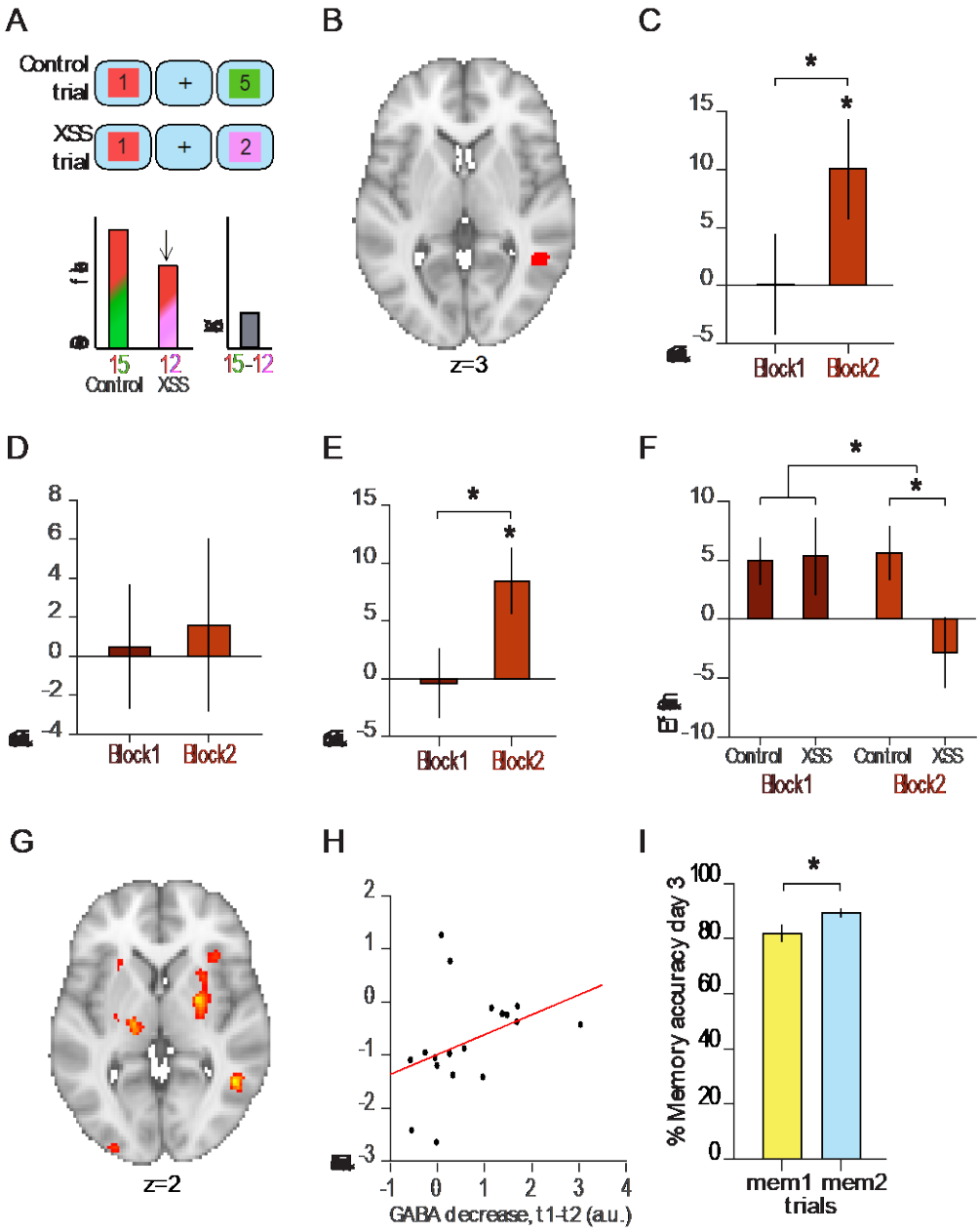


Supplementary Figure 5.2 Example of fMRI partial volume and RSA models with additional RSA analyses

A-C To increase SNR in brain regions for which we had strong prior hypotheses, we restricted the fMRI sequence to a partial volume, thus allowing for an increase in the number of measurements acquired due to shorter TR. The partial volume covered occipital and temporal cortices. Here, brain regions included in the partial volumes of all participants are shown for the (A) sagittal, (B) coronal and (C) axial plane. Orientation: neurological. **D-F** Model representational dissimilarity matrices (RDMs) used to assess evidence for: **D** Within versus between memory dissimilarity, with results shown in Figure 5.3B; **E** Between memory dissimilarity for stimuli that change their relational position across memory 1 and 2 (i.e. stimuli 3 and 6) relative to all other stimuli (1,2,4,5,7), with results shown in Figure 5.3C; **F** Within versus between stimulus dissimilarity, with results shown in Supplementary Figure 5.2I. **G** The RDM for each participant was correlated with a model RDM shown in Supplementary Figure 5.2D to test evidence for pattern separation of hippocampal representations by memory. All trials with either a 3 or 6 stimulus were excluded from the analysis. Across participants, significant representational similarity within memory 1/2 versus between memory 1 and 2 was observed (Wilcoxon sign rank test: Z_{23}=2.46, p=0.014). Note: the dissimilarity of a trial to itself was excluded from the analysis. 'a.u.' refers to 'arbitrary units'. **H** ROI shown in Figure 5.4D, centered on peak tDCS electrode location (see *Methods*). Orientation: neurological. **I** For each participant, we assessed the dissimilarity in hippocampal activity patterns between- versus within-stimulus exemplars in memory 1 and in memory 2 (e.g. [1 to 2, 1 to 3, 1 to 4, etc] minus [1 to 1, 2 to 2, 3 to 3, etc]), and observed a significant positive difference within the anterior LOC ROI shown in H (between – within stimulus exemplars: Wilcoxon sign rank test: Z_{23}=2.75, p=0.003). 'a.u.' refers to 'arbitrary units'. Note: the dissimilarity of a trial to itself was excluded from the analysis.

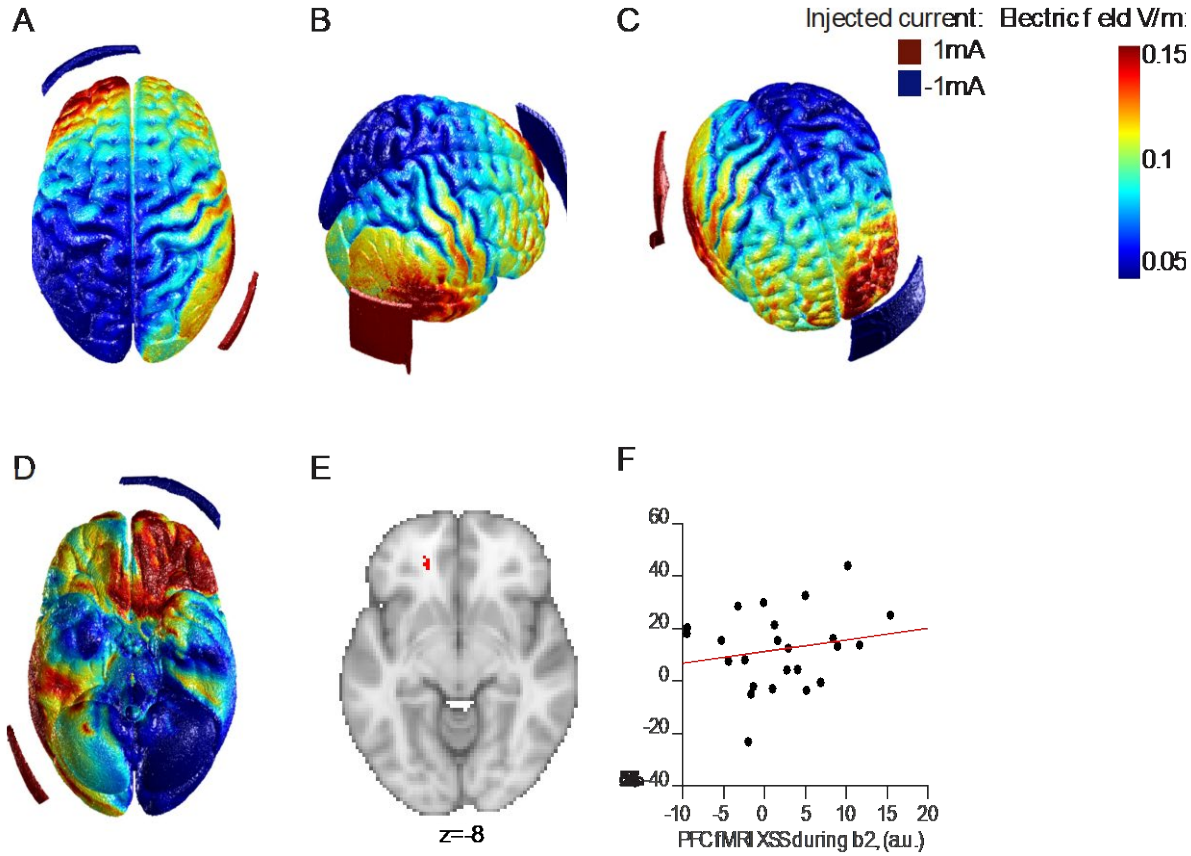


Supplementary Figure 5.3 MRS spectra, subject inclusion, and additional analyses
A Average spectra for all participants' data included in MRS analysis (shown: mean ±SEM). Chemical shifts of the three GABA peaks are indicated using green stars. **B** Average spectra for those participants who were rejected from the MRS analysis (shown: mean ±SEM). Data from these participants were noisy, and had lipid contamination in the region 1.9-0.5ppm (i.e. in the region of the lowest GABA peak). This resulted in either inestimable or highly unreliable GABA estimates. Chemical shifts of the three GABA peaks are indicated using green stars. **C** Relative to participants included in the MRS analysis (shown in A), those participants rejected from the MRS analysis (shown in B) had broader linewidth, estimated using full-width at half maximum (FWHM) using LCModel (two-sample t-test: t_{24}=3.73, p=0.001) (shown: mean ±SEM). **D** There was no significant change in the concentration of Creatine across the 3 MRS measurements ('Before tDCS' – 'During', t_{19}=0.26, p=0.799; 'Before tDCS' – 'Post task' t_{19}=1.37, p=0.186; shown: mean ±SEM). **E-F** By assessing metabolite concentrations relative to Creatine and by comparing the concentration of GABA/glutamate between two time points in a within subject manner, our analyses (Figure 5.4E-F) controlled for variation in voxel tissue and CSF in the MRS voxel used across subjects and thus mitigated the need for partial volume correction. To check that the proportion of CSF in the MRS voxel did not affect water signal referencing in LCModel we reassessed the change in the concentration of GABA and glutamate across the 3 MRS sessions, after accounting for the proportion of CSF. We show that our results shown in Figure 5.4E-F remain unchanged: **E** A significant reduction in relative GABA was observed during tDCS ('Before tDCS' – 'During tDCS', t19= 2.24, p=0.019, shown: mean ±SEM). **F** A significant increase in relative glutamate was observed after the second scan task ('Post-task' – 'Before tDCS', t_{19}= 2.56, p=0.019, shown: mean ±SEM).



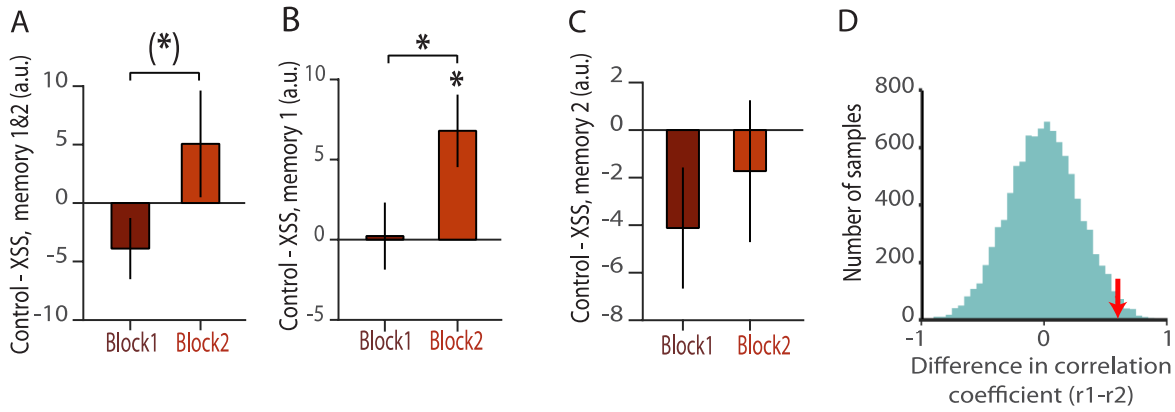
Supplementary Figure 5.4 Replication of previous findings (Barron et al., 2016a)
'XSS' indicates cross-stimulus suppression; 'b' indicates block of fMRI acquisition; 't' indicates timepoint of MRS acquisition. **A** When participants performed the scan task in E/I imbalance, we predicted an increase in cross-stimulus suppression ('XSS') on trials where pairs of directly associated stimuli were presented (e.g. stimuli 1 and 2), relative to trials where pairs of unassociated stimuli were presented ('Control', e.g. stimuli 1 and 5). The difference between 'Control' and 'XSS' trials could be indexed using the BOLD signal to provide a measure of cross-stimulus suppression. **B** To test replication of our previously published result (Barron et al., 2016a) (shown in Figure 5.4B), we used an independently defined ROI in aLOC, taken from our previous dataset (Figure 5.4B thresholded at $p<0.01$, see Methods). Orientation: neurological. **C** For directly associated stimuli across both memory 1 and 2, extracted parameter estimates (shown: mean ±SEM) revealed a significant increase in fMRI cross-stimulus suppression during brain stimulation, and significant fMRI cross-stimulus suppression during brain stimulation (within the ROI shown in B: 'Control' – 'XSS' for Block 2 – Block 1: $t_{23}=1.73$, $p=0.049$; for Block 2: $t_{23}=2.31$, $p=0.015$). Thus, replicating our previous findings (Barron et al., 2016a), these results show that reducing GABAergic tone increases repetition suppression between associated stimuli. **D** Between directly associated stimuli in memory 1, extracted parameter estimates (shown: mean ±SEM) revealed no significant change in fMRI cross-stimulus suppression during brain stimulation (within the ROI shown in B: 'Control' – 'XSS' for Block 2 – Block 1: $t_{23}=0.23$, $p=0.823$). **E** Between directly associated stimuli in memory 2, extracted parameter estimates (shown: mean ±SEM) revealed a significant increase in fMRI cross-stimulus suppression during brain stimulation, and significant fMRI cross-stimulus suppression during brain stimulation (within the ROI shown in B: 'Control' – 'XSS'

for Block 2 – Block 1: t_{23}=2.29, p=0.016; Block 2: t_{23}=2.92, p=0.004). **F** Extracted parameter estimates from E split into the 'Control' and 'XSS' conditions, as described in A (shown: mean ±SEM). **G** T-statistic map for cross-stimulus suppression between directly associated stimuli in memory 2 during Block 2 shown in E-F, thresholded at p<0.01 uncorrected for visualisation. Significant cross-stimulus suppression for directly associated elements of memory 2 could also be observed in Block 2 within a 10mm radius sphere defined around the peak tDCS electrode location for all participants (Figure 5.4D, see Methods) (t_{23}=4.17, p=0.010, peak-level FWE corrected using SVC). Orientation: neurological. **H** There was a significant positive correlation between the change in relative GABA ('Before tDCS' – 'During tDCS') and the increase in fMRI cross-stimulus suppression (Block 2- Block 1) observed in the ROI shown in Figure 5.4D, averaged across both memory 1 and 2 (Spearman correlation: r_{17}=0.52, p=0.028, after accounting for changes in glutamate, see Methods). Thus, the decrease in relative GABA during application of anodal tDCS positively predicted the increase in cross-stimulus suppression between directly associated stimuli in memory 1 and 2. 'b' indicates block for fMRI acquisition as shown in Figure 5.1F. 't' indicates 'timepoint' of MRS measurement acquisition, as shown in Figure 5.1F. **I** Across memory 1 and memory 2 (Figure 5.1B-C), 3 of the 7 associations remained the same: those between stimuli 1 and 2, 1 and 7 and 4 and 5. Memory accuracy on the surprise memory test (shown: mean ±SEM) for these stable associations was significantly different between memory 1 (day 1, yellow) and memory 2 (day 2, blue), with higher memory accuracy when participants recalled these associations in the more recent memory 2 (paired t-test: t_{25}=2.16, p=0.040).



Supplementary Figure 5.5 Simulated effect of tDCS across the brain and effect of cathodal stimulation on left prefrontal cortex (PFC)

A-D Using open source software, ROAST, we used a current model to estimate the electric field induced by tDCS across neocortex for our electrode configuration. To simulate the effect of the anodal electrode, a 1mA current was injected into the right anterior LOC. To simulate the effect of the cathodal electrode, a -1mA current was injected into the left prefrontal cortex. This stimulation configuration gave rise to an elevated electric field (V/m) under both the anodal and cathodal electrodes. **A** Dorsal surface. **B** Right hand hemisphere, including region under the anodal electrode. **C** Anterior view, showing the region under cathodal electrode. **D** Ventral surface. **E-F** Our cross-suppression contrast for memory interference did not show any suprathreshold voxels (defined as $p<0.001$, uncorrected) underneath the cathodal electrode. This suggests that the effects reported from anterior LOC (Figure 5.5) cannot be explained by downstream effects of cathodal stimulation near to left PFC. However, to check that *subthreshold* fMRI effects in PFC (defined as $p<0.01$ uncorrected) cannot explain the reported effects in anterior LOC (Figure 5.5) we assessed the relationship between these two brain regions. **E** ROI in PFC, defined from a contrast for our cross-stimulus suppression index for memory interference, thresholded at $p<0.01$ uncorrected. Orientation: neurological. **F** The relationship between PFC and anterior LOC: parameter estimates were extracted from PFC (ROI shown in *E*) and from anterior LOC (peak average tDCS electrode location shown in Figure 5.4C) during Block 2 of the fMRI scan task and the cross-stimulus suppression index for memory interference was assessed. The cross-stimulus suppression in PFC did not predict the measure for memory interference in anterior LOC ($r_{23}=0.10$, $p=0.626$). This suggests that cross-stimulus suppression effects reported in anterior LOC (Figure 5.5) cannot be explained by downstream of cathodal stimulation near to left PFC. 'XSS' indicates cross-stimulus suppression; '*b*' indicates block for fMRI acquisition, as shown in Figure 5.1F.



Supplementary Figure 5.6 Memory interference effects in memory 1 and 2
A Cross-stimulus suppression was used to index the change in memory interference before and during application of tDCS, as shown in Figure 5.5 (shown: mean ±SEM). Across both memory 1 and 2 there was a trend towards an increase in the cross-stimulus suppression index for memory interference (paired t-test: t_{23}=1.79, p=0.087). **B** As shown in Figure 5.5E: within an ROI defined from the peak average tDCS electrode location shown in Figure 4.4C, extracted parameter estimates for memory 1 (shown: mean ±SEM) revealed a significant increase in the fMRI cross-stimulus suppression measure for memory interference ('Control' – 'XSS', as shown in Figure 5.5A) from Block 1 to 2 and during Block 2 alone ('Control' – 'XSS' for Block 2 – Block 1: t_{23}=3.05, p=0.006; 'Control' – 'XSS' for Block 2: t_{23}=3.00, p=0.006). **C** Unlike for memory 1 shown in B and Figure 5.5E-F, no significant change in the fMRI cross-stimulus suppression measure for memory interference was observed for memory 2 (t_{23}=0.57, p=0.573, shown: mean ±SEM). **D** During Block 1 but not Block 2 of the scan task hippocampal BOLD predicted subsequent behavioural performance on the surprise memory test (Figure 5.2H and Figure 5.6C). To assess the significance of the difference in correlation coefficient, a null distribution of 10,000 samples was estimated using a permutation test (see Methods). Here, the null distribution can be observed in green and the difference in correlation indicated by the red arrow.

5.8 Supplementary Tables

Supplementary Table 5.1 Average concentration of all metabolites measured using MRS
For each metabolite, the concentration was measured relative to Creatine and then averaged across participants (shown: mean, ± SEM). As reported in the main text, a significant decrease in the concentration of GABA was observed during tDCS, and a significant increase in the concentration of glutamate observed after the second task block (Figure 5.4E-F). Of the other metabolites measured (n=17), only one showed a significant change in concentration across the three MRS measurements: the concentration of aspartate significantly decreased after the second task block ('before tDCS' – 'post-task', t_{19}=4.29, $p<0.001$).

Metabolite	Before tDCS	During tDCS	Post-task
Alanine	1.39 ± 0.42	1.52 ± 0.44	1.31 ± 0.45
Ascorbate	1.11 ± 0.17	1.29 ± 0.18	1.22 ± 0.17
Aspartate	3.18 ± 0.34	2.78 ± 0.34	2.26 ± 0.27
Glycerophosphorylcholine	0.93 ± 0.07	0.86 ± 0.08	0.94 ± 0.09
Phosphorylcholine	0.75 ± 0.06	0.83 ± 0.06	0.76 ± 0.07
Creatine	4.96 ± 0.14	4.97 ± 0.15	4.91 ± 0.14
Phosphocreatine	3.04 ± 0.14	3.03 ± 0.15	3.09 ± 0.14
GABA	4.77 ± 0.34	4.18 ± 0.45	4.53 ± 0.39
Glucose	2.22 ± 0.18	2.31 ± 0.21	2.31 ± 0.15
Glutamine	7.30 ± 0.37	7.36 ± 0.42	7.04 ± 0.36
Glutamate	9.26 ± 0.26	9.39 ± 0.22	9.75 ± 0.26
Glutathione	0.89 ± 0.08	0.77 ± 0.11	0.79 ± 0.10
Inositol	7.27 ± 0.15	7.19 ± 0.14	7.19 ± 0.16
Lactate	1.10 ± 0.20	1.18 ± 0.15	0.89 ± 0.14
N-acetylaspartate (NAA)	13.25 ± 0.37	13.19 ± 0.34	13.28 ± 0.39
N-acetylaspartylglutamate (NAAG)	2.34 ± 0.11	2.29 ± 0.12	2.31 ± 0.10
Phosphoethanolamine	1.91 ± 0.17	1.97 ± 0.15	1.95 ± 0.19
Scyllo-Inositol	0.18 ± 0.03	0.17 ± 0.03	0.17 ± 0.03
Taurine	0.65 ± 0.08	0.53 ± 0.08	0.51 ± 0.07

CHAPTER 6

MEMORY RECALL INVOLVES A TRANSIENT BREAK IN EXCITATORY-INHIBITORY BALANCE

The previous chapter demonstrated that memories can be released from a dormant state by lowering cortical inhibitory activity. In this chapter I investigate whether disinhibition is indeed associated with memory recall, and how this cortical mechanism for memory reactivation relates to activity in hippocampus. I hypothesised that the hippocampus reactivates memories by modulating E/I balance in neocortex. To test this idea, I simultaneously assessed brain activity using fMRI and the concentration of glutamate and GABA using time-resolved fMRS while participants recalled previously learned auditory-visual associations. The results revealed that the ratio of glutamate to GABA in visual neocortex increased during successful recall, which was predicted by activity in the hippocampus. This suggests that the hippocampus reactivates memories stored in neocortex by temporarily increasing excitatory activity to release memories from inhibitory control.

* This chapter is published in the following article:

Koolschijn RS†, Shpektor A†, Clarke WT, Ip IB, Dupret D, Emir UE, Barron HC (2021). Memory recall involves a transient break in excitatory-inhibitory balance. *Elife* 10:e70071.

† indicates equal contribution

6.1 Abstract

The brain has a remarkable capacity to acquire and store memories that can later be selectively recalled. These processes are supported by the hippocampus which is thought to index memory recall by reinstating information stored across distributed neocortical circuits. However, the mechanism that supports this interaction remains unclear. Here, in humans, we show that recall of a visual cue from a paired associate is accompanied by a transient increase in the ratio between glutamate and GABA in visual cortex. Moreover, these excitatory-inhibitory fluctuations are predicted by activity in the hippocampus. These data suggest the hippocampus gates memory recall by indexing information stored across neocortical circuits using a disinhibitory mechanism.

5.2 Introduction

Memories are thought to be stored across sparse and distributed neuronal ensembles in the brain (Buzsáki, 2010; Josselyn and Tonegawa, 2020). During memory recall, activity across these neuronal ensembles is selectively reinstated to recover enduring representations of the past. This reinstatement is thought to be mediated by the hippocampus, a brain region important for learning and memory (Squire, 1992). Anatomically, the hippocampus sits at the apex of a cortical sensory processing hierarchy (Felleman and Essen, 1991) where inputs received by sensory cortices reach the hippocampus via the entorhinal cortex and other relay regions, which in turn make widespread cortico-cortical connections that project the hippocampal output back to neocortex (Witter, 1993; Witter et al., 1989). This reciprocal anatomical connectivity equips the hippocampus with the necessary architecture to coordinate activity in neocortex. The hippocampus may therefore be considered to provide a 'memory index', or summary sketch, for information stored across distributed cortical circuits (Goode et al., 2020; Teyler and DiScenna, 1985; Teyler and Rudy, 2007). Consistent with this view, during memory recall hippocampal reinstatement predicts subsequent neocortical reinstatement (Pacheco Estefan et al., 2019; Tanaka et al., 2014).

However, the mechanism that allows the hippocampus to coordinate reinstatement across distributed neocortical circuits remains unclear. One possibility is that the hippocampus shapes computations performed by neocortical circuits by modulating the dynamic interplay between excitation and inhibition (E/I). At the cellular level, tight coupling between neocortical excitation and inhibition can be observed during both sensory stimulation and spontaneous neural activity (Haider et al., 2006; McCormick et al., 2004; Okun and Lampl, 2008; Wehr and Zador, 2003). This phenomenon has led to the physiological concept of E/I balance, where, following changes in

excitability, synaptic strength, current or overall network activity returns to a stable set point via negative feedback (Field et al., 2020). Evidence in humans, animal models and theoretical models together suggest E/I balance is maintained to hold memories in a silent and dormant state (Barron et al., 2016a; Froemke et al., 2007; Vallentin et al., 2016; Vogels et al., 2011), thus protecting memories from interference caused by new learning (Koolschijn et al., 2019; Kuchibhotla et al., 2017). During recall, however, E/I balance must be transiently disturbed if memories are to be released from inhibitory control.

Here we predict memory recall involves a transient break in E/I balance, opening a window to release memories from the blanket of inhibition before network stability is re-established. Moreover, we predict this transient break in neocortical E/I balance is mediated by activity in the hippocampus. To test these predictions, here we implemented a new imaging sequence in humans that combines functional Magnetic Resonance Imaging (fMRI) with functional Magnetic Resonance Spectroscopy (fMRS) (Ip et al., 2019, 2017). This sequence provides an opportunity to monitor activity in the hippocampus with fMRI while simultaneously measuring time-resolved fluctuations in neocortical glutamate and GABA using fMRS.

MRS provides a unique tool to quantify the concentration of different neural metabolites (De Graaf, 2019; Mangia et al., 2012), including glutamate and GABA, the principle excitatory and inhibitory neurotransmitters in the brain. MRS cannot dissociate between neurotransmitter and metabolic pools of glutamate and GABA (Bak et al., 2006; Magistretti and Allaman, 2015), however meaningful interpretation of MRS nevertheless derives from a major body of work showing an approximately 1:1 relationship between the rate of glutamine-glutamate cycling, which is necessary for glutamate and GABA synthesis, and neuronal oxidative glucose consumption, which indirectly supports neurotransmitter release among other processes (Rothman et al., 2003; Shen et al., 1999; Sibson et al., 1998). Therefore, while measures of E/I balance vary in both definition and granularity, MRS can provide a non-invasive marker for physiologically relevant excitation and inhibition at a coarse spatiotemporal scale. Correspondingly, MRS-derived glutamate and GABA reported during learning and memory paradigms show remarkable consistency with findings reported at the physiological level in animals (Barron et al., 2016a; Castro-Alamancos et al., 1995; Floyer-Lea et al., 2006; Froemke et al., 2007; Kolasinski et al., 2019; Lunghi et al., 2015; Trepel and Racine, 2000; Vallentin et al., 2016).

Using the combined fMRI-fMRS sequence, here we implemented a task designed to engage hippocampal-dependent recall of a visual cue. During memory recall, we report a transient increase in the ratio between MRS-derived glutamate and GABA in neocortex which is selectively predicted

by the Blood-Oxygen-Level-Dependent (BOLD) signal in the hippocampus. These findings suggest the hippocampus coordinates memory recall by transiently perturbing neocortical E/I balance to release memories stored across distributed neural circuits.

6.3 Methods

6.3.1 Participants

22 healthy human volunteers were included in the study (mean age of 22.8 ± 0.74 years, 4 males). All experiments were approved by the University of Oxford ethics committee (reference number R43594/RE001). All participants gave informed written consent. For one participant, we were unable to collect combined fMRI-fMRS data due to time constraints during scanning. Two participants were excluded from the fMRI and fMRS analyses due to technical difficulties which resulted in the auditory cues not being fully audible during the inference test. Notably, there was no significant effect of sex on either behavioural performance or MRS measures of glu/GABA ratio during the inference test.

6.3.2 Virtual reality environment

The virtual reality (VR) environment was coded using Unity 5.5.4f1 software (Unity Technologies, CA United States) (Figure 6.1A). The VR environment was designed to simulate an open field environment previously used to investigate memory and inference in mice (Barron et al., 2020). Within the VR environment participants were exposed to a range of different sensory stimuli, in accordance with the three-stage inference task described below.

The environment included a square-walled room with no roof (Figure 6.1A). To help evoke the experience of 3D space and aid orientation within the VR environment, each wall of the environment was distinguished by colour (dark green, light green, dark grey or light grey), illumination (two walls were illuminated while the other two were in shadow) and by the presence of permanent visual cues. The permanent visual cues included clouds in the sky, a vertical black stripe in the middle of the light green wall, a horizontal black strip across the light grey wall, and a wooden box situated in one corner of the environment. A first-person perspective was implemented and participants could control their movement through the virtual space using the keyboard arrows (2D translational motion) and the mouse-pad (head tilt). Movement through the environment elicited the sound of footsteps. Within the VR environment participants were

exposed to a range of different sensory stimuli, in accordance with the three-stage inference task described below.

6.3.3 Three-stage inference task

In the VR environment (Figure 6.1A) humans performed an inference task (Figure 6.1B). The rationale for using an inference task to assess mechanisms responsible for associative memory was three-fold. First, evidence in both humans (Barron et al., 2020; Koster et al., 2018) and mice (Barron et al., 2020) shows that performance on this inference task requires associative memory recall. Second, in rodents, inference, but not first-order associative recall, is hippocampal dependent (Barron et al., 2020; Bunsey and Eichenbaum, 1996; DeVito et al., 2010), thus providing an opportunity to investigate hippocampal dependent associative memory recall. Third, the task can be deployed across humans and rodents, which may allow future investigation of the cellular mechanisms that underlie non-invasive measures reported here.

The task was adapted from associative inference and sensory preconditioning tasks described elsewhere (Barron et al., 2020; Brogden, 1939; Preston and Eichenbaum, 2013) and involved 3 stages performed across 3 consecutive days, respectively (Figure 6.1B,C). The first and second stages were performed outside the scanner while the third stage was performed inside the scanner (Figure 6.1C). At the start of the experiment the pairings between auditory, visual and outcome cues were randomly assigned for each participant.

On day 1, participants performed the 'associative learning' stage (Figure 6.1B), during which participants were required to learn at least 40 (out of 80 total) auditory-visual associations via mere exposure. In total, there were 4 visual cues, each associated with 20 different auditory cues. Auditory cues constituted 80 different complex sounds (e.g. natural sounds or those produced by musical instruments) that were played over headphones. Visual cues constituted 4 different unique patterned panels which could appear on the walls of the environment (Figure 6.1A,B). To control for potential spatial confounds, two of the visual cues were always presented on the same wall, the assignment of which was randomised for each participant. The two remaining visual cues were 'nomadic', meaning that with each presentation they were randomly assigned to one of the four walls.

Training during the associative learning stage occurred within the VR environment as described previously (Barron et al., 2020). In brief, on each trial of the 'associative learning' an auditory and visual cue were presented serially and contiguously: 8 s auditory cue followed by 8 s of the associated visual cue, followed by an ITI of 5 s (Supplementary Figure 6.1A). Learning of auditory-

visual associations was monitored outside the VR environment, using an associative learning test coded in Matlab 2016b using Psychtoolbox (version 3.0.13). On each trial of the associative learning test, 1 auditory cue was presented, followed by presentation of 4 different visual cues (Supplementary Figure 6.1B). Participants were instructed to select the visual cue associated with the auditory cue using a button press response within 3 s, and only at the end of the test were participants given feedback on their average performance. Training on the associative learning stage was terminated only when participants reached >50% accuracy on the associative learning test when all 80 auditory cues were included, each presented 3 times (Supplementary Figure 6.1E). Those participants that failed to reach >50% accuracy (n=3) did not proceed to day 2 and were thus not included in the experiment.

On day 2, participants performed the 'conditioning' stage (Figure 6.1B), during which they learned that two of the four visual cues (set 1) predicted delivery of a rewarding outcome (virtual silver coin, as above) on 80% of trials, while the other two visual cues (set 2) predicted delivery of a neutral outcome (virtual wood-chip, as above) on 100% of trials. Training during the conditioning stage occurred within the VR environment and on each trial, participants were presented with a visual cue (8 s) followed by outcome delivery to a wooden box (available for 6 s) situated in the corner of the environment (Supplementary Figure 6.1C). The inter-trial interval (ITI) was 2 s. To harvest the value of a virtual silver coin (monetary reward later converted to 20 pence per coin) or woodchip (no value, 0 pence), participants were required to collect the coin or wood-chip from the wooden box. The cumulative total value of harvested reward was displayed in the upper left corner of the computer screen.

Learning during conditioning was monitored using a conditioning test coded in Matlab 2016b using Psychtoolbox (version 3.0.13). On each trial of the conditioning test, participants were presented with a still image of a visual cue before being asked to indicate the probability of reward using a number line (Supplementary Figure 6.1D). Participants were given 3 s to respond and were only given feedback on their average performance at the end of the test. Participants were required to repeat the VR conditioning training and conditioning test until they performed the test with 100% accuracy (Supplementary Figure 6.1F).

Finally, on day 3, participants first repeated the conditioning test. Participants then entered the 7T MRI scanner and performed the 'inference test' (Figure 6.1B-D), together with a subset of conditioning trials (Supplementary Figure 6.10A) (see *fMRI-fMRS scan task* below). Immediately after exiting the scanner, participants were given a surprise associative test to assess which auditory-visual associations they remembered (Figure 6.2A). The memory test was equivalent to

the test performed on day 1 during the associative learning (Supplementary Figure 6.1B), with 3 trials for each auditory stimulus. Performance on auditory-visual associations was categorised as correct if participants scored 3/3 for a given auditory-visual pair. Performance on auditory-visual associations was categorised as incorrect if participants scored 0/3 or 1/3 for a given auditory-visual pair (i.e. no different from chance). Trials where participants scored 2/3 were not categorised as either correct or incorrect due to their ambiguity. The behavioural performance measured on the post-scan associative test (Figure 6.2A) was a more sensitive measure of memory accuracy than behavioural performance measured during the inference test alone, with a lower chance level (associative test: 4 choice options with 1.6% chance level for correct response across 3 repeats; inference test: 2 options with 50% chance level for correct response across 1 repeat). For this reason, performance on the inference test during the scan was assessed post-hoc using performance from both the inference test and the post-scan associative test (see *Trial categorisation during the inference test*, Figure 6.3B).

6.3.4 fMRI-fMRS scan task

The inference test was incorporated into the fMRI-fMRS scan task. This provided an opportunity to measure neural responses to associative memory recall required for inferential judgements. The scan task included two different trial types: inference test trials (Figure 6.1D) and conditioning trials (Supplementary Figure 6.10A). For both types of trial participants viewed a short video taken from the VR training environment. The videos were presented via a computer monitor and projected onto a screen inside the scanner bore. On each trial the duration of the video was determined using a truncated gamma distribution with mean of 7 s, minimum of 4 s and maximum of 14 s. During the inference test trials, the video of the VR environment was accompanied by an auditory cue, played over MR compatible headphones (S14 inset earphones, Sensimetrics). Visual cues were not displayed during these trials: the auditory cues were presented in isolation. At the end of the video, participants were presented with a question asking: 'Would you like to look in the box?', with the options 'yes' or 'no' (Figure 6.1D). Importantly, as described above, outcomes (rewarding or neutral) were delivered to the wooden box during the conditioning stage. Participants were required to make a response within 3 s using an MR compatible button box and their right index or middle fingers. No feedback was given. To infer the appropriate outcome participants were instructed to use the learned structure of the task. After each trial (inference or conditioning) a cross was presented in the centre of the screen during an inter-trial interval of varying length, determined using a truncated gamma distribution (mean of 2.7 s,

minimum of 1.4 s, maximum of 10 s) Trials were categorised as 'correctly inferred' if participants pressed 'yes' in response to auditory cues indirectly associated with a rewarding outcome, or pressed 'no' in response to auditory cues indirectly associated with a neutral outcome (Figure 6.1E). The inference test provided an opportunity to investigate memory recall: to infer the correct outcome participants needed to recall the appropriate visual cue associated with the auditory cue (Figure 6.2C-D).

Conditioning trials were interleaved with inference test trials to minimise extinction effects. During conditioning trials, the video of the VR environment orientated towards a visual stimulus displayed on one of the four walls (Supplementary Figure 6.10A). At the end of the video, participants were presented with a still image of the associated outcome for that visual cue (Supplementary Figure 6.10A).

To control for potential confounding effects of space, each video during the inference test involved a trajectory constrained to a 1/16 quadrant of the VR environment, evenly distributed across the different auditory cues. Across conditioning trials, each visual cue was presented 16 times, once in each possible spatial quadrant. Moreover, the videos were not related to the background for the relevant visual cue. Allocation of the videos to each trial was randomised separately for each participant to ensure no consistent biases. The fMRI-fMRS scan task was evenly divided across 2 scan blocks, each of which lasted 15 minutes. The fMRI-fMRS scan task was then repeated (2 more scan blocks) using a higher quality multiband fMRI sequence (reported elsewhere (Barron et al., 2020).

6.3.5 fMRI-fMRS data acquisition

The fMRI-fMRS scan task was performed inside a 7 Tesla Magnetom MRI scanner (Siemens) using a 1-channel transmit and a 32-channel receive phased-array head coil (Nova Medical Inc., USA) at the Wellcome Centre for Integrative Neuroimaging (University of Oxford). Current 7T radio-frequency (RF) coil designs suffer from B_1^+ inhomogeneity. To overcome this, we positioned two 110 × 110 × 5 mm^3 Barium Titanate dielectric pads (4:1 ratio of $BaTiO_3$:D_2O, relative permittivity around 300) over occipital lobe, causing a "hotspot" in the proximal B_1^+ distribution at the expense of distal regions (Brink and Webb, 2014). For each participant, a T1-weighted structural image was acquired to inform placement of the MRS voxel in visual cortex, and to correct for geometric distortions and perform co-registration between EPIs, consisting of 176 0.7 mm axial slices, in-plane resolution of 0.7 × 0.7 mm^2, TR = 2.2 s, TE = 2.96 ms, and field of view

= 224 mm. For each participant, a field map with dual echo-time images was also acquired (TE1 = 4.08 ms, TE2 = 5.1 ms, whole-brain coverage, voxel size 2 × 2 × 2 mm^3).

Figure 6.3A shows a diagram of the combined fMRI-fMRS sequence, based on a sequence developed by Hess et al. (Hess et al., 2011), and previously used to compare the BOLD signal in V1 with measures of glutamate (Ip et al., 2019, 2017). In the same TR of 4.105 s, BOLD-fMRI (3D EPI, resolution 2.3 × 2.3 × 2.2 mm^3; flip angle=5°, repetition time TR_{epi}= 59 ms, TE=29 ms, field of view 200 mm, 32 slices) and fMRS data (2 × 2 × 2 cm^3 voxel positioned in the occipital lobe, centered along the midline and the calcarine sulcus) were acquired. The TR was increased for 4 participants where a higher transmit voltage was required, resulting in a TR between 4.7-5.9 s. fMRS data were acquired using short-echo-time semi-localisation by adiabatic selective refocusing (semi-LASER) pulse sequence (TE=36 ms, TR_{mrs}=4 s) with VAPOR water suppression and outer volume suppression (Öz and Tkáč, 2011). A delay between fMRI and fMRS acquisition (250 ms) was inserted to minimise potential eddy current effects from the EPI read-out (Hess et al., 2011). Compared to an uncombined contemporary MR sequences (e.g. multiband EPI and semi-LASER MRS), the fMRS was of comparable quality, while the quality of the fMRI component was compromised. The quality of MRS data was assessed during set-up scans and acquisition. Several criteria were considered: we monitored the level of noise in the data and the residual water signal to check for stability and to ensure that the size of the water peak was well below the level of most observable metabolites. During set-up scans, the position of the MRS voxel was adjusted if necessary. Data quality was further assessed during data analysis, using the metrics reported in Supplementary Figure 6.5 and listed in Supplementary Table 6.5. On average, 457 fMRS spectra were acquired over the two scanning blocks (SD: 35.62).

In addition to the fMRI-fMRS sequence acquisition, an additional set of fMRI data (reported elsewhere (Barron et al., 2020) and not shown here) was acquired using a multiband EPI sequence (50 1.5 mm thick transverse slices with 1.5 mm gap, in-plane resolution of 1.5 × 1.5 mm^2, TR=1.512 s, TE= 20 ms, flip angle = 85°, field of view 192 mm, and multi-band acceleration factor of 2). To increase SNR in brain regions for which we had prior hypotheses, both the fMRI sequences were restricted to partial brain coverage (Figure 6.3A, covering the occipital and temporal lobes) to shorten the EPI TR, thus acquiring more measurements.

6.3.6 Trial categorisation during the inference test

Trials during the inference test were categorised into two conditions, 'remembered' and 'forgotten' (Figure 6.3B). Given participants were required to make a binary 'yes'/'no' response in

the inference test, chance was at 50%. To ensure our findings could not merely be explained by false positives, we categorised trials during the inference test into 'remembered' and 'forgotten' using a conservative approach, to ensure we could confidently identify when the associated visual cue was recalled in response to the auditory cue. To this end, we controlled for false positives in the inference test by filtering behavioural performance on the inference test using behavioural performance recorded on the post-scan associative memory test (where correct recall of auditory-visual associations has chance level of 1.6%). Thus, trials where participants made both the correct inference during the inference test *and* indicated the correct auditory-visual association during the post-scan associative test were classified as 'remembered'. Trials where participants made *either* the incorrect inference during the inference test *or* the incorrect auditory-visual association during the post-scan associative test were classified as 'forgotten'. Notably, this approach effectively eliminates trials where participants were guessing during the inference test, thus providing a conservative estimate of trials where the subjects remember the auditory-visual associations.

6.3.7 fMRS metabolite quantification and analysis

For each scan run, fMRS data from 19 subjects was preprocessed separately in MRspa, a semi-automated MATLAB routine (https://www.cmrr.umn.edu/downloads/mrspa/). The unsuppressed water signal acquired from the same VOI was used to remove residual eddy current effects and combine individual coil spectra. Spectra were corrected for frequency and phase variations induced by participants' motion, and the residual water component was removed using Hankel Lanczos Singular Value Decomposition (HLSVD). For each participant, spectra from all blocks were frequency aligned to account for frequency differences between blocks. The parameters implemented for fMRS acquisition are summarised in Supplementary Table 6.8, according to standards proposed by the 'MRS Experts Working Group' (Lin et al., 2021).

Spectra were then analysed in an event-related manner. For each participant, the preprocessed spectra were first assigned to the tone/question/ITI periods by aligning the time stamps for the spectra to the time stamps for each event recorded during the inference task. Then, spectra acquired within the tone/question/ITI periods were selected for analysis. Next, these selected spectra were separated into two categories according to task performance, 'remembered' or 'forgotten' (Figure 6.3B, see *Trial categorisation during the inference test*), before being analysed using LCModel. Participants (n=1 for the 'Question' period only) with less than 8 spectra for either the 'remembered' or 'forgotten' conditions were excluded from the fMRS analysis. Notably, previous studies report minimal change in test-retest CoVs when going from 8 to 16 spectra (Terpstra et

al., 2016). Metabolite concentrations for the average 'remembered' and the average 'forgotten' spectrum were quantified in turn using LCModel (Provencher, 1993) within the chemical shift range 0.5 to 4.2 ppm. The concentration of each metabolite was assessed relative to the concentration of total Creatine (Creatine + phosphocreatine, tCr), thus providing effective control for variation in voxel tissue and cerebral spinal fluid (CSF) in the fMRS voxel used across participants. Within LCModel, metabolite estimates were not constrained by priors that assume the GABA estimate remains fixed relative to more abundant neurochemicals. As a result, relative to default settings, the GABA estimates were higher and the dynamic range of GABA was not assumed to be fixed (Supplementary Figure 6.4). Estimates were normalised to tCr and multiplied by 8, as per convention. A basis set containing stimulated model spectra of alanine (Ala), aspartate (Asp), ascorbate/vitamin C (Asc), glycerophosphocholine (GPC), phosphocholine (PCho), creatine (Cr), phosphocreatine (PCr), GABA, glucose (Glc), glutamine (Gln), glutamate (Glu), glutathione (GSH), myo-inositol (myo-Ins), Lactate, N-acetylaspartate (NAA), N-acetylaspartylglutamate (NAAG), phosphoethanolamine (PE), scyllo-inositol (scyllo-Ins), taurine (Tau) and experimentally measured macromolecules was used.

Changes in the relative concentration of glutamate and GABA between 'remembered' and 'forgotten' conditions were evaluated together with 'glu/GABA ratio' which we defined as the ratio of glutamate to GABA (Shibata et al., 2017). We defined the change in glutamate, GABA and glu/GABA for 'remembered' vs 'forgotten' trials as a ratio, as follows:

$$Glutamate\ ratio = 100 \times \left(\frac{Glu_{remem} - Glu_{forgot}}{Glu_{forgot}}\right)$$

$$GABA\ ratio = 100 \times \left(\frac{GABA_{remem} - GABA_{forgot}}{GABA_{forgot}}\right)$$

$$glu/GABA\ ratio = 100 \times \left(\frac{Glu_{remem}/GABA_{remem} - Glu_{forgot}/GABA_{forgot}}{Glu_{forgot}/GABA_{forgot}}\right)$$

Where Glu and GABA represent the ratio of glutamate and GABA to total Creatine, respectively, during the tone/question/ITI period of 'remembered' or 'forgotten' trials. This ratio effectively controls for variation in voxel tissue and CSF fraction in the MRS voxel used across participants.

Further, to control for differences in the number of 'remembered' and 'forgotten' spectra, we compared the group mean difference between 'remembered' and 'forgotten' trials against a null distribution generated by permuting the trial labels while preserving differences in number of trials for each participant. On each of 5000 permutations, the condition labels ('remembered', 'forgotten') were shuffled for each participant using MATLAB's random number generator. The relative metabolite concentrations for each condition were then estimated in LCModel and the difference between conditions computed. The group mean for each permutation was then added to the null distribution. The difference between 'remembered' and 'forgotten' conditions derived from the unshuffled data was then compared against the null distribution generated from the shuffled data (Figure 6.4D-F; Supplementary Figure 6.6-9).

To control for a false positive result and to provide further evidence for the reported change in glu/GABA ratio between 'remembered' and 'forgotten' trials, we generated a sampling-error distribution computed from 10,000 bootstrapped resamples of glu/GABA ratio (Efron, 2000). We visualised the effect size of the relative measure using a Data Analysis with Bootstrap-coupled ESTimation (DABEST) plot (Ho et al., 2019) (Figure 6.4F).

6.3.8 MRS Monte Carlo simulations

To assess the relative sensitivity to detecting changes in GABA in our data, we used Monte Carlo simulations to generate synthetic spectra. The average observed spectrum (across participants) was used as an input to Monte Carlo simulations (Clarke et al., 2021). For each set of conditions we generated 2000 simulated spectra, with the SNR and linewidth of the simulated data matched to the SNR and linewidth observed in the data. The T2 values are assumed to be the same between conditions since we did not see any differences in FWHM between 'remembered' and 'forgotten' conditions in the *in vivo* data (Supplementary Figure 6.6B). The output of the simulations was then analysed in LCModel to quantify GABA and glutamate.

These simulated spectra were used for two types of analyses. First, the simulated spectra were used to test the likelihood of observing the measured change in GABA between 'remembered' and 'forgotten' conditions by chance (Figure 6.4G). Second, the simulated spectra were used to assess the effect of synthetically imposing changes in GABA, both with and without constraining GABA relative to the concentration of other more abundant neurochemicals ('constraints on' versus 'constraints off') (Supplementary Figure 6.4). The imposed changes in GABA were the following multiples of the observed difference in GABA (between 'remembered' and 'forgotten'): 0, ±0.5, ±1, ±2. To assess the sensitivity of the 'constraints on' and 'constraints off' LCModel

settings to changes in imposed GABA at different SNRs, we compared the slope for each setting. Slopes were determined by fitting a GLM to the imposed vs measured GABA concentration in the simulated data. To quantify the difference in slope between categories (Supplementary Figure 6.4C: 'constraints on' vs. 'constraints off'; Supplementary Figure 6.4D: differences in SNR) we randomly sampled n=18 simulated spectra for each imposed change in GABA, for each condition of interest (e.g. 'constraints on' vs. 'constraints off'; SNR 125% vs. SNR 75% etc.). Over 500 sets of size n=126 (i.e. n=18 simulated spectra for each imposed change in GABA), we estimated the power to reject the null hypothesis of equal slopes between conditions. Similarly, using a parametric statistical approach, over 500 sets of size n=126 simulated spectra we estimated the t-statistic to reject the null hypothesis of equal slopes between conditions.

6.3.9 fMRI preprocessing and GLMs

Preprocessing of MRI data was carried out using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). First, the anterior commissure was set to the origin in the anatomical images and in the first volume of each fMRI block, with equivalent transformations applied to all other images within the same block. Second, to account for magnetic field inhomogeneities, images were corrected for signal bias, realigned to the first volume, corrected for distortion using field maps, normalised to a standard EPI template. To remove low frequency noise from the pre-processed data, a high-pass filter was applied to the data using SPM12's default settings. For each participant and for each scanning block, the resulting fMRI data was analysed in an event-related manner using a general linear model (GLM). The GLM was applied to data from both scan task blocks. In addition to the explanatory variables (EVs) of interest (described below), 6 additional scan-to-scan motion parameters produced during realignment were included in the GLM as nuisance regressors to account for motion-related artefacts in each task block. The output of the first-level analysis was then smoothed using a 5-mm full-width at half maximum Gaussian kernel before being entered into a second level analysis. The sensitivity of our analysis pipeline to detecting stimulus evoked BOLD activity patterns benefitted from applying the first-level GLM to unsmoothed data and only including smoothing prior to the second level analysis (Supplementary Figure 6.2). Across participants, data quality was assessed after each step in the pre-processing of the data. One participant was excluded from the fMRI analyses as co-registration between the epi and structural scans failed due to insufficient quality of fMRI data. Notably, this participant was not the same as the participant excluded from the fMRS analyses, where less than 8 spectra in either the 'remembered' or 'forgotten' conditions occurred for n=1 participants (see above).

For the first-level analyses, three different GLMs were used. Each GLM included 15 EVs per block. In the first GLM, the first 8 EVs accounted for the question period in the inference test, divided according to performance of the subject ('remembered' or 'forgotten', see *Trial categorisation during the inference test*), before being further divided according to the 4 possible visual cues to which the auditory cues were associated. The next 4 EVs accounted for presentation of the visual cue during the video of all conditioning trials, divided according to the 4 different visual cues. The final 3 EVs accounted for presentation of the auditory cue during the video in all inference test trials, the question period in all remaining inference test trials (i.e. trials not categorised as 'remembered' or 'forgotten'), and the presentation of the outcome in all conditioning trials. To decorrelate the EVs modelling the auditory and visual cues from those EVs modelling the question and outcome, respectively, the duration of events within EVs modelling the auditory and visual cues was set using a box-car function to 4 s, i.e. the minimum duration of the video. The duration of events within EVs modelling the question/outcome were set to the duration of the question/outcome. All EVs were then convolved with the hemodynamic response function.

In the second and third GLMs, the same EVs were included, however the first 8 EVs accounted for the auditory cue period in the inference test (second GLM), or the inter-trial interval in the inference test (third GLM). In both cases, the EVs were divided according to performance of the subject ('remembered' or 'forgotten'), as in the first GLM.

6.3.10 Univariate fMRI analysis and statistics

Using the output of the GLMs we assessed the difference in the univariate BOLD response between 'remembered' and 'forgotten' trials during the inference test (as defined in Figure 6.3B, *Trial categorisation during the inference test*). The contrast of interest therefore involved contrasting EVs [1:4] ('remembered') with EVs [5:8] ('forgotten'), using the first GLM (see above). The resulting contrast images ('remembered'-'forgotten') for all participants were entered into a second-level random effects 'group' analysis. We set the cluster-defining threshold to $p<0.01$ uncorrected before using whole-brain family wise error (FWE) to correct for multiple comparisons, with the significance level defined as $p<0.05$ (Figure 6.3C, Supplementary Table 6.4).

6.3.11 Assessing the relationship between fMRI and fMRS

To assess the relationship between event-related hippocampal BOLD signal and event-related fMRS measures from V1, we used an anatomical ROI for the hippocampus (Figure 6.5A).

Capitalising on variance across participants, the relationship between the BOLD signal for 'remembered'-'forgotten' within this ROI was compared with equivalent changes in glutamate, GABA and glu/GABA ratio using a Spearman rank correlation. To assess the selectivity of these effects to the recall period (question) during the inference test, control analyses were performed using the output of the second and third GLMs, together with equivalent measures of glutamate, GABA and glu/GABA ratio (Supplementary Figure 6.11).

Next, to assess the relationship between fMRS and the BOLD signal across the entire imaged brain volume, we repeated the second-level random effects 'group' analysis using the output of the first GLM, but now included group-level covariates for the change in glutamate and GABA for 'remembered'-'forgotten' (i.e. Figure 6.4A), along with 2 'nuisance' regressors that accounted for unwanted variance attributed to differences in age and sex. To identify brain regions where the BOLD signal for 'remembered':'forgotten' predicted changes in glu/GABA ratio, we contrasted the explanatory variables on the covariates for glutamate and GABA (glutamate – GABA) to generate a single contrast to test statistical significance. We set the cluster-defining threshold to $p<0.01$ uncorrected before using whole-brain family wise error (FWE) to correct for multiple comparisons, with the significance level defined as $p<0.05$ (Figure 6.5C, Supplementary Table 6.7).

To visualise the time course of fMRS across the inference test trials, we estimated a moving average, where each time bin constituted a 2.5 s time window shifted by 0.5 s in each iteration (Figure 6.4C, Figure 6.5D, Supplementary Figure 6.11A-B). By incorporating a random jitter in each trial of the fMRI-fMRS scan task, the temporal relationship between MRS spectra acquisition and the inference test trials varied. Thus, across participants and across trials, MRS spectra were acquired in all possible 2.5 s time bins of the inference test trial, achieving a higher temporal resolution than the TR of 4 s (Supplementary Figure 6.3). To ensure each time bin contained a similar number of spectra, those bins at the tail end of the jitter (final 3 time bins during the video and the final 2 time bins of the ITI) were enlarged to include broader time windows (>2.5 s). For each participant the 'remembered' and 'forgotten' spectrum were then calculated for each time bin, and the ratio estimated to give a measure of 'remembered':'forgotten' for both glutamate and GABA. For each time bin, data for a given participant was only included if the participant had more than 8 spectra for both 'remembered' and 'forgotten' conditions for that bin. The number of participants per time bin therefore varied (mean: n=17.47; range: n=12 to n=19).

To visualise the time course of data acquired using fMRI, for each participant, and for each time bin during the inference test trial, the time course of the preprocessed BOLD signal was extracted from the hippocampal ROI (Figure 6.5A) and from two control ROIs defined using a

12 mm sphere within our partial epi volume (Figure 6.3A). The first control region was positioned at the junction between parietal and occipital cortex ('parietal-occipital cortex') while the second control region was positioned within the brainstem (Supplementary Figure 6.11A-B). For each ROI, the obtained signal for each trial was resampled using a resolution of 400 ms and regressed against an explanatory variable indicating those trials that were 'remembered'. To control for differences in baseline BOLD at the start of the trial, we also included a 'nuisance' explanatory variable indicating whether the previous trial was 'remembered'. We then plotted the normalised averaged fMRI regression coefficient for 'remembered' vs 'forgotten', using the time bins defined for the fMRS moving average (described above) (Figure 6.5D; Supplementary Figure 6.11A-B).

6.4 Results

6.4.1 Task design and behaviour

To investigate the neuronal mechanisms that support memory recall we designed a three-stage inference task. This task has previously been shown to involve associative memory recall in humans (Barron et al., 2020; Koster et al., 2018) and mice (Barron et al., 2020). Unlike some forms of associative recall, previous lesion and optogenetic studies in rodents demonstrate that associative recall required for inference is a hippocampal dependent process (Barron et al., 2020; Bunsey and Eichenbaum, 1996; DeVito et al., 2010). Thus, the inference task provides an opportunity to investigate whether activity in the hippocampus mediates dynamic changes in neocortical excitation and inhibition during memory recall.

The inference task was performed in virtual reality (VR) (Figure 6.1A), an immersive and highly controlled 3D environment that has the potential to benefit from cross-species comparisons in the future (Barron et al., 2020). The inference task was performed across three days and included three stages (Figure 6.1B). In the first stage of the task participants learned up to 80 auditory-visual associations ('associative learning', day 1; Figure 6.1B, Supplementary Figure 6.1). In the second stage, which occurred approximately 24 hours later, each visual cue was paired with either a rewarding (set 1, monetary reward) or neutral outcome (set 2, woodchip) delivered to a wooden box in the corner of the VR environment ('conditioning', day 2; Figure 6.1A-B, Supplementary Figure 6.1). Auditory cues were never paired with an outcome, providing an opportunity to assess evidence for an inferred relationship between these indirectly related stimuli.

Accordingly, in the third stage of the task we presented auditory cues in isolation, without visual cues or outcomes, and we measured evidence for inference from the auditory cues to the

appropriate outcome ('inference test', day 3; Figure 6.1B). Participants performed the inference test during an MRI scan (Figure 6.1C-D). On each trial of the inference test participants were presented with an auditory cue, before being asked if they would like to look in the wooden box ('yes' or 'no') where they had previously found the outcomes during the conditioning stage. Participants' responses depended upon whether they inferred the indirectly associated outcome to be rewarding or neutral. On trials where the auditory cue was associated with a visual cue paired with a rewarding outcome (set 1 cues), participants were expected to select 'yes' if they inferred the relevant outcome (Figure 6.1E). On trials where the auditory cue was instead associated with a visual cue paired with a neutral outcome (set 2 cues), participants were expected to select 'no' if they inferred the relevant outcome (Figure 6.1E). We thus categorised trials during the inference test as 'correctly inferred' if participants selected 'yes' when the auditory cue was indirectly associated with rewarding outcome or 'no' when the auditory cue was indirectly associated with a neutral outcome.

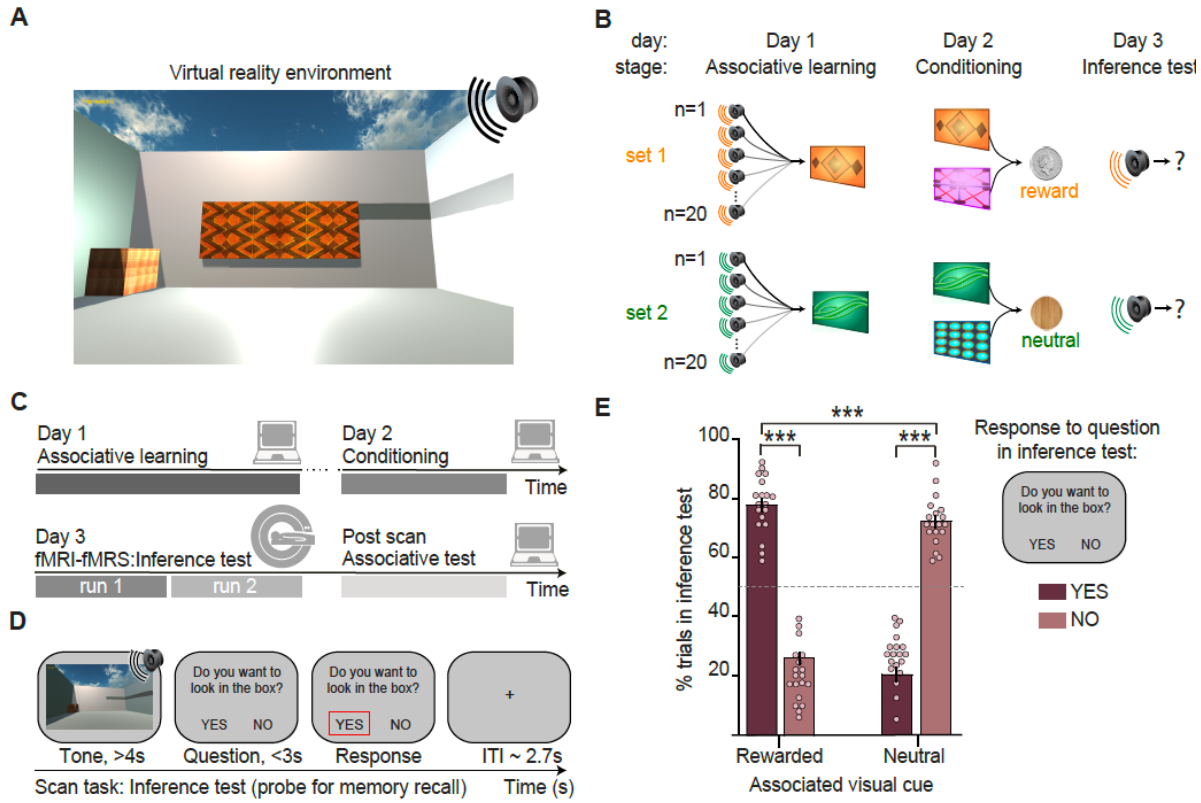


Figure 6.1 Inference task design and behavioural performance

A The inference task was performed within a virtual-reality environment. **B** Three-stage inference task designed to investigate hippocampal-dependent associative memory recall. First, participants learned to associate auditory cues with visual cues ('associative learning' stage, day 1), where four different visual cues were each associated with 20 auditory cues. Second, participants learned to associate visual cues with an outcome delivered to a wooden box in the corner of the virtual-reality environment ('conditioning' stage, day 2). Two visual cues predicted a rewarding outcome (set 1, monetary coin) while the other two predicted a neutral outcome (set 2, woodchip). Third, the auditory cues were played in isolation and we assessed participants' ability to infer the relevant outcome by recalling the intermediary visual cue ('inference test', day 3). **C** Schematic: training and testing protocol. The inference test was performed inside the 7T MRI scanner. After exiting the scanner, participants were given a surprise post-scan associative test to directly assess participants' memory for auditory-visual associations learned on day 1. **D** Example

inference test trial performed inside the scanner. For each auditory cue, participants were required to infer the indirectly associated outcome by recalling the relevant auditory-visual association. To indicate whether participants inferred the outcome to be rewarding or neutral, on each trial participants pressed 'yes' or 'no' in response to a question asking, 'Would you like to look in the box?', referring to the box where the outcome cues were delivered during conditioning. **E** Behaviour during the inference test revealed a significant interaction between the response to auditory cues in set 1 and 2 ("rewarded" and "neutral") and whether the participants indicated whether or not they wanted to look in the box ("yes" and "no") (two-way ANOVA, $F_{1,72} = 630.99$, $p < 0.001$). Tukey's post hoc test showed participants pressed "yes" more often for auditory cues in set 1 ($p < 0.001$), and "no" more often for auditory cues in set 2 ($p < 0.001$). This inferential behaviour was observed despite participants never experiencing the outcomes in response to the auditory cues. For the purpose of the analyses reported in Figures 6.2-6.5, trials where participants pressed "yes" for auditory cues in set 1, or "no" for auditory cues in set 2 were categorised as "correctly inferred" trials. *** indicates $p<0.001$.

Previous studies using this task show that in trials where participants infer the correct outcome, the associated visual cue that links the auditory cue and outcome is reinstated in the hippocampus and visual cortex (Barron et al., 2020). Consistent with these previous findings, here we show that participants make the correct inference only if they can later recall the relevant auditory-visual association during a surprise post-scan associative test (Figure 6.1C, Figure 6.2A-C). Indeed, performance on the post-scan associative test that assessed memory for auditory-visual associations learned on day 1, predicted performance on the inference test (Figure 6.2D). The inference task thus provides a suitable paradigm to investigate the neural mechanisms that support associative recall, in this case for auditory-visual associations.

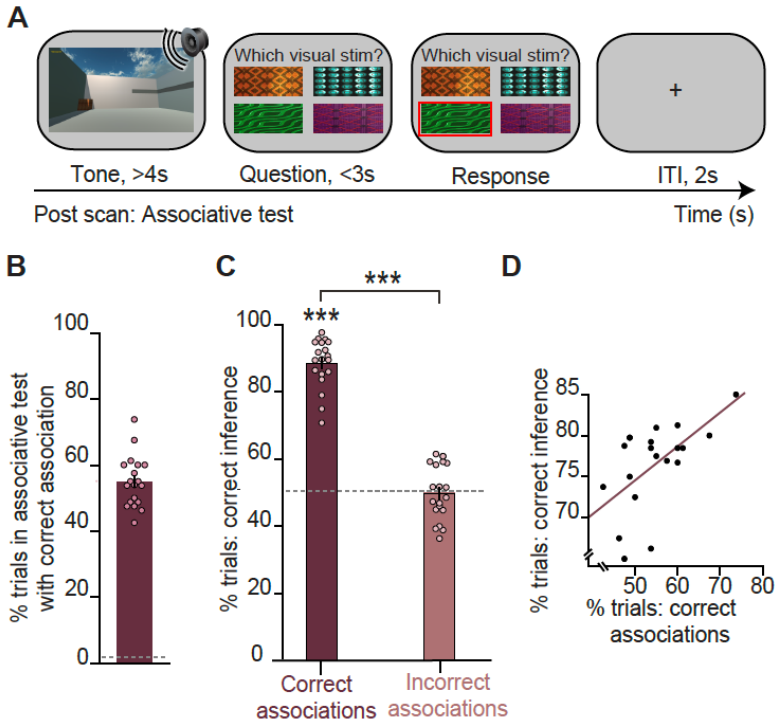


Figure 6.2 Behavioural performance in the inference test is predicted by performance on an associative test
A Example trial from the surprise post-scan associative test which directly tested participants' memory for auditory-visual associations learned on day 1. **B** During the post-scan associative test participants remembered 55% of the auditory-visual associations (54.8 ±1.78%; mean ±SEM), significantly above chance as indicated by the dotted line ($t_{18}=29.96$, $p<0.001$). **C** In response to auditory cues during the inference test (Figure 6.1D), participants successfully inferred the appropriate outcome (Figure 6.1E) on trials where they could later recall the correct auditory-visual association in the post-scan associative test ('correct association': $t_{18}=22.91$, $p<0.001$; 'incorrect association': $t_{18}=0.09$,

p=0.925; 'correct association' – 'incorrect association': t_{18}=16.21, p<0.001; dotted line indicates chance). **D** Across participants, behavioural performance on the inference test was predicted by behavioural performance on the post-scan associative test (r_{17}=0.57, p=0.010). Notably, there was no significant effect of sex on behavioural performance (Supplementary Table 6.1). *** indicates p<0.001.

6.4.2 Neural signatures of associative memory recall during inference

To investigate neural signatures of associative memory recall during the inference test we implemented a novel imaging sequence (Ip et al., 2019, 2017) which enabled interleaved acquisition of near-whole brain fMRI together with fMRS in primary visual cortex (V1) (Figure 6.3A).

The fMRI-fMRS imaging sequence (Figure 6.3A) provided a means to simultaneously measure both hemodynamic and neurochemical changes in an event-related manner. By incorporating a temporal jitter in each trial of the experimental paradigm (Figure 6.1D), the relationship between data acquisition and the experimental paradigm varied on a trial-by-trial basis (Supplementary Figure 6.3). Therefore, across trials it was possible to effectively assess data at a higher temporal resolution than that given by a TR of 4 s.

In the inference test participants were required to make a binary 'yes'/'no' response, with chance at 50%. To exclude trials where participants guessed, we classified trials as 'remembered' or 'forgotten' using a conservative approach. We filtered trials during the inference test post-hoc using participants' behavioural performance from the subsequent post-scan associative test (Figure 6.2B). Trials where participants made both the correct inference (inference test; chance 50%) and indicated the correct auditory-visual associations (associative test; chance 1.6%) were classified as 'remembered'. Trials where participants made either the incorrect inference (inference test) or indicated an incorrect auditory-visual association (associative test) were classified as 'forgotten' (Figure 6.3B, Supplementary Table 6.2, *Methods*). Neural signatures acquired during the 'forgotten' trials thus provided a condition- and stimulus-matched control for data acquired during the 'remembered' trials. Notably, this approach to categorising trials during the inference test controlled for false positives in the inference test, providing a conservative estimate of trials where participants remembered the auditory-visual associations. Notably, there was no significant difference between the number of trials in set 1 (rewarding) versus set 2 (neutral) for the 'remembered' and 'forgotten' conditions (memory x set, two-way ANOVA: $F_{(1,68)}$=0.67, p=0.424; Supplementary Table 6.3).

Using the fMRI data from the interleaved sequence, we first identified brain regions modulated by recall of a visual cue during the inference test (Figure 6.1D). Consistent with previous research investigating associative recall of visual cues (Horner et al., 2015; Wimmer and Shohamy, 2012) and data acquired using the same task (Barron et al., 2020), we observed a significant increase in

BOLD signal in both the hippocampus and visual cortex on 'remembered' versus 'forgotten' trials (Figure 6.3C; Supplementary Figure 6.2).

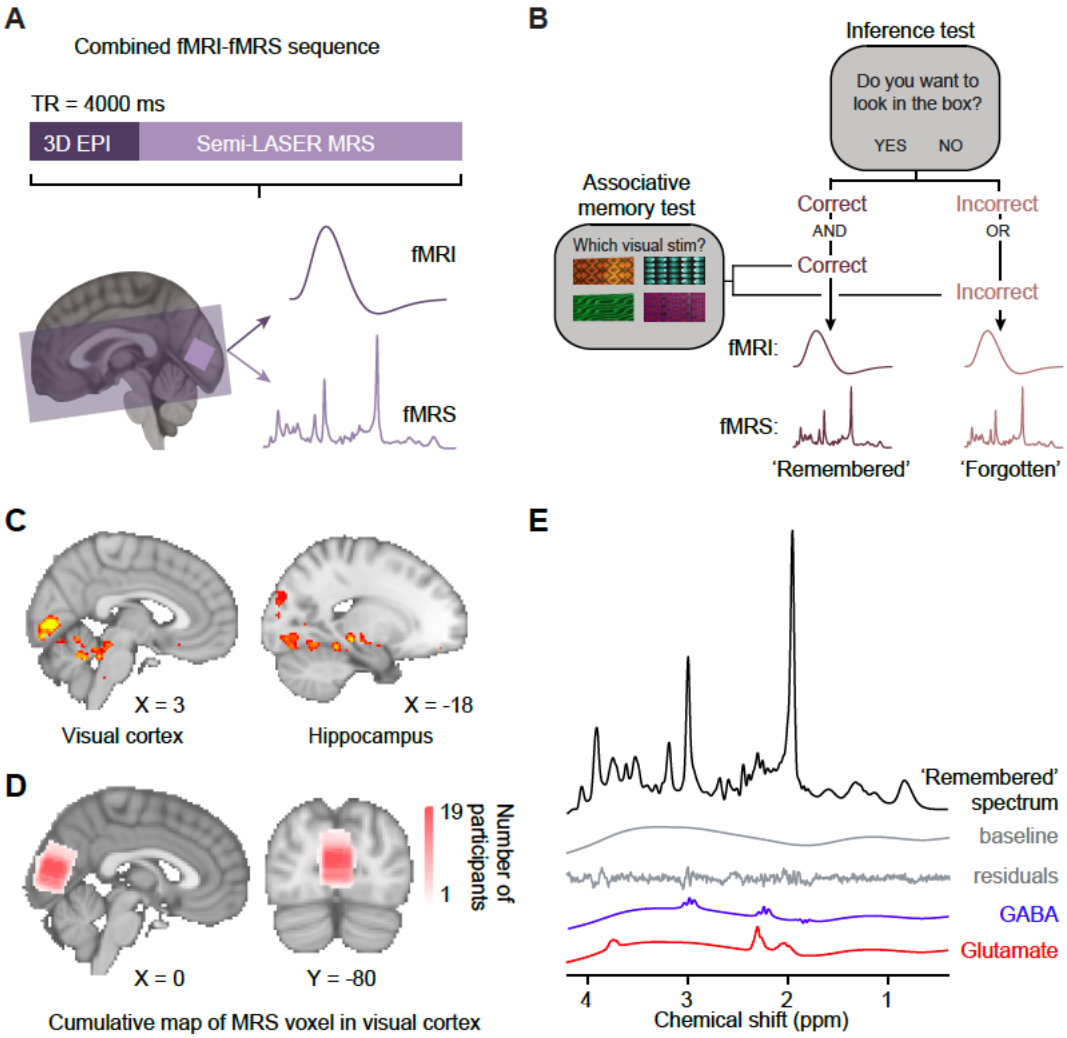


Figure 6.3 Using fMRI-fMRS data to assess changes in BOLD signal and glu/GABA ratio during the inference test

A 7T MRI sequence. 3D BOLD echo planar imaging (3D-EPI) and semi-LASER MR-spectroscopy were acquired in the same TR. The MRS voxel was positioned in V1 (light-purple) and the EPI slice coverage included occipital and temporal lobes (dark-purple). **B** Schematic showing how trials during the inference test were categorised into 'remembered' and 'forgotten'. Trials were categorised as 'remembered' if participants correctly inferred the appropriate outcome during the inference test *and* subsequently indicated the correct auditory-visual association in the post-scan associative test. Trials were categorised as 'forgotten' if participants incorrectly inferred the appropriate outcome during the inference test *or* indicated the incorrect auditory-visual association in the post-scan associative memory test. Notably, by using performance on the associative test to filter trials on the inference test, this approach helped eliminate false positive and false negative responses during the inference test where chance was otherwise at 50% (Figure 6.2C). This conservative approach thus provides a more reliable measure of associative memory recall during inference. **C** During the question period in the inference test (Figure 6.1C-D), BOLD signal in the visual cortex and the hippocampus was significantly higher for 'remembered' versus 'forgotten' auditory cues ('remembered' – 'forgotten', visual cortex: t_{17}=5.92, p<0.001; hippocampus: t_{17}=4.33, p=0.017; whole-volume FWE-corrected; together with regions listed in Supplementary Table 6.4; Montreal Neurological Institute (MNI) coordinates). **D** Anatomical location of 2x2x2 cm^3 MRS voxel positioned in V1. Cumulative map across participants; MNI coordinates. **E** Representative MRS spectrum from 'remembered' trials in the inference test for an example participant.

6.4.3 Dynamic increase in the ratio between glutamate and GABA in visual cortex during recall

We then asked whether associative memory recall of a visual cue is accompanied by changes in the ratio between glutamate and GABA ('glu/GABA ratio', see *Methods*) in visual cortex. We chose this region of interest because recalling a visual cue is known to involve reinstating cortical representations in visual cortex (Bosch et al., 2014; Wheeler et al., 2000), including during inference as verified with an independent fMRI data set using the same task (Barron et al., 2020). Using the interleaved fMRS data acquired in primary visual cortex (V1) (Figure 6.3A,D), we quantified the concentration of glutamate and GABA normalised to total Creatine (tCr) in an event-related manner (Figure 6.3B, E). Notably, to assess dynamic changes in GABA, in the metabolite fitting procedure it was not appropriate to employ default settings used to detect static estimates of GABA. Importantly, these default settings constrain values of GABA relative to more stable metabolites, a process that effectively limits the dynamic range of GABA (Supplementary Figure 6.4). Instead, here we use unconstrained GABA estimates (see *Methods*): while this approach leads to GABA estimates that are higher than values normalised by the concentration of more stable metabolites, critically, dynamic changes in GABA can be detected (Supplementary Figure 6.4).

We used MRS-derived measures of glutamate and GABA to estimate changes in glu/GABA ratio (Shibata et al., 2017). During associative memory recall in the inference test we observed an increase in glu/GABA ratio in V1 when comparing 'remembered' versus 'forgotten' trials (Figure 6.4A-C). Standard quality metrics indicated that our data quality was reliable over the course of the acquisition (Supplementary Figure 6.5, Supplementary Table 6.5). To control for any biases introduced by differences in the number of 'remembered' versus 'forgotten' trials (Supplementary Table 6.6), we compared the group mean metabolite change against a null distribution generated by permuting the identity labels ('remembered' or 'forgotten') assigned to each trial. This analysis again revealed a significant increase in glu/GABA ratio during memory recall, together with a significant decrease in GABA (Figure 6.4D-F).

These findings cannot be explained by differences in data quality measures between the 'remembered' and 'forgotten' conditions (Supplementary Figure 6.6). In addition, the reported change in glu/GABA ratio was still observed when categorising trials into 'remembered' and 'forgotten' using performance on the inference task alone, a less conservative approach (Supplementary Figure 6.7). The increase in glu/GABA ratio was not observed during periods immediately before or after recall (Figure 6.4A-B; Supplementary Figure 6.8). Moreover, no effect between 'remembered' and 'forgotten' was observed in NAA, a neurometabolite that has

overlapping peaks with GABA but is found at higher concentration (Supplementary Figure 6.9). Notably, the observed within-subject, task-specific changes in neurochemistry were obscured when assessing the relationship between glutamate and GABA across subjects (r_{17}=0.191, p=0.433; after regressing out sex and age: r_{17}=0.205 p=0.400), consistent with previous findings (Rideaux, 2021). Thus, we propose the reported transient increase in neocortical glu/GABA ratio reflects a mechanism for associative memory recall.

As an additional control, we assessed changes in glu/GABA ratio during a subset of conditioning trials (Supplementary Figure 6.10A) that were interleaved with the inference test trials during the MRI scan and shared the same temporal structure. Importantly, previous work suggest performance on conditioning trials is not hippocampal-dependent (Barron et al., 2020). During the conditioning trials, we observed no change in glu/GABA ratio during presentation of the visual cue or outcome, relative to the ITI period (Supplementary Figure 6.10B-C).

We note that our MRS sequence does not use editing techniques which exploit known *J*-coupling relationships to separate signals deriving from low concentration metabolites, such as GABA, from stronger, overlapping signals (Mullins et al., 2014). Instead, we implemented an MRS sequence without editing while taking advantage of the benefits associated with using a short TE. To further assess the sensitivity of our approach to detecting dynamic changes in GABA across task conditions, we used Monte Carlo simulations to generate MRS spectra while preserving the observed noise in our data. Using these simulations we show that the observed difference in GABA between 'remembered' and 'forgotten' conditions is significant from a null distribution that would be expected by chance (Figure 6.4G).

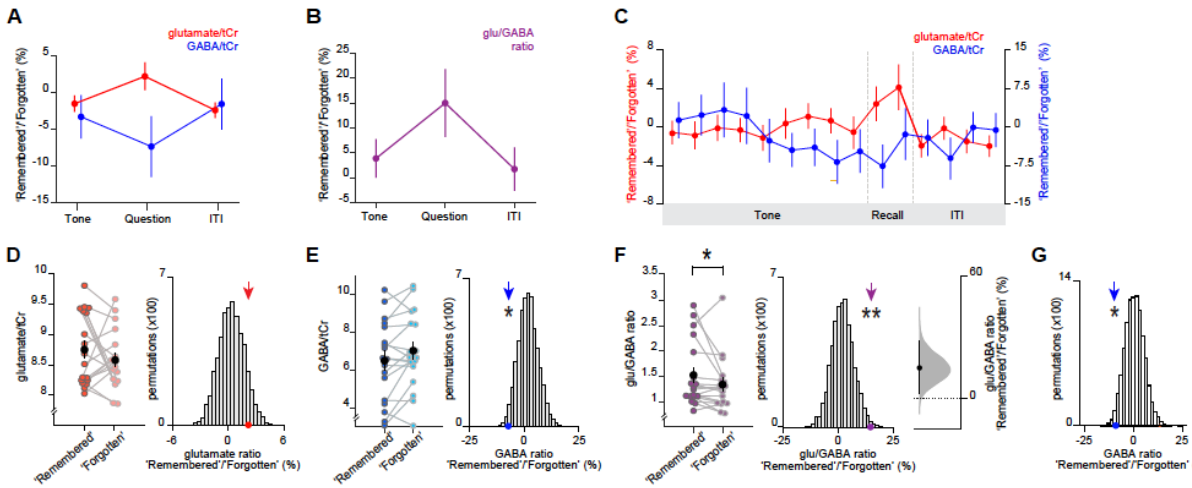


Figure 6.4 Memory recall and inference involves a transient break in glu/GABA ratio
A-B During the question period of the inference test trials (up to 3 s), glu/GABA ratio significantly increased during 'remembered' versus 'forgotten' trials ('remembered':'forgotten', glu/GABA ratio: t_{17}=2.21, p=0.041). One participant (n=1) was excluded from this analysis, where there were less than 8 spectra for either the 'remembered' or

'forgotten' conditions during the question period (see *Methods*). This break in glu/GABA ratio was not observed during the 'tone' (~7 s) or 'ITI' (~2.7 s) periods ('Tone', glu/GABA ratio: t_{18}=1.01, p=0.325; 'ITI', glu/GABA ratio: t_{18}=0.40, p=0.692). Note that glutamate:tCR and GABA:tCr concentrations have been multiplied by 8 as per convention. To detect dynamic changes in glu/GABA we chose not to use LCModel's default settings which assume the dynamic range of GABA is fixed (see *Methods*; Supplementary Figure 6.4). **C** Moving average showing glutamate:tCr and GABA:tCr for the ratio of 'remembered' to 'forgotten' trials during the inference test. Each point represents a 2.5 s time bin (mean ± SEM). By incorporating a random jitter in the behavioural paradigm, MRS spectra across all trials and all participants were acquired in all possible 2.5 s time bins of the inference test trial (Supplementary Figure 6.3), thus achieving a higher temporal resolution than the TR of 4 s (see *Methods*). **D-F** Left: The metabolite values and glu/GABA ratio during the question period for 'remembered' and 'forgotten' trials (mean ± SEM). **D, E** Right, **F** Middle: Comparing the mean ratio of 'remembered' to 'forgotten' (coloured arrows) against null distributions generated by permuting the trial labels to control for any potential biases in the analyses. Relative to the null distributions, GABA significantly decreased while glu/GABA ratio significantly increased (glutamate:tCr: p=0.089; GABA:tCr: p=0.014; glu/GABA ratio: p=0.007). Note: To detect dynamic changes in GABA, it was not appropriate to normalise GABA estimates relative to the concentration of more stable metabolites (see *Methods*). Consequently, GABA values are higher than those generated using default settings in LCModel which are optimised for detecting static estimates (Supplementary Figure 6.4). **F** Right panel: full sampling-error curve for glu/GABA ratio estimated using bootstrap-coupled estimation (DABEST) plot (Ho et al., 2019). The 95% confidence interval is non-overlapping with zero (p=0.017). Black dot, mean; black tick, 95% confidence interval; filled-curve, sampling-error distribution. **G** The average measured spectra were used as an input to Monte Carlo simulations, to generate simulated spectra with the level of noise matched to the observed data (see Supplementary Figure 6.4). Using this simulated data, we established a null distribution for the difference between pairs of 'remembered' and 'forgotten' spectra that would be expected by chance (i.e. when the condition labels are shuffled). Relative to this null distribution, the observed GABA ratio measured *in vivo* (shown in ***e***) was significant (p=0.019). * indicates p<0.05, ** indicates p<0.01.

6.4.4 A hippocampal index for fluctuations in neocortical ratio between glutamate and GABA

We next asked which brain regions coordinate this transient break in neocortical glu/GABA ratio during memory recall. The hippocampus is a promising candidate, given this brain region supports memory (L.R. Squire, 1992) and shows activity modulation during the inference test (Figure 6.3C). To test this possibility, we took advantage of our simultaneous fMRI-fMRS acquisition (Figure 6.3A). We hypothesised that the increase in hippocampal BOLD signal observed during recall (Figure 6.3C) should predict the increase in glu/GABA ratio observed in V1 (Figure 6.4B, F). In line with this prediction, across participants the hippocampal BOLD signal negatively predicted the relative concentration of GABA, and positively predicted the increase in glu/GABA ratio in V1 ('remembered' versus 'forgotten' trials; Figure 6.5A-B). This relationship between the BOLD signal and glu/GABA ratio was not observed in two control regions of interest (ROIs) (Supplementary Figure 6.11A-B). Furthermore, across the imaged brain volume (Figure 6.3A), only the hippocampus significantly predicted the increase in V1 glu/GABA ratio on 'remembered' versus 'forgotten' trials (Figure 6.5C). Finally, this relationship between the hippocampus and glu/GABA ratio was specific to the recall period during the inference test (Figure 6.5D, Supplementary Figure 6.11C-D).

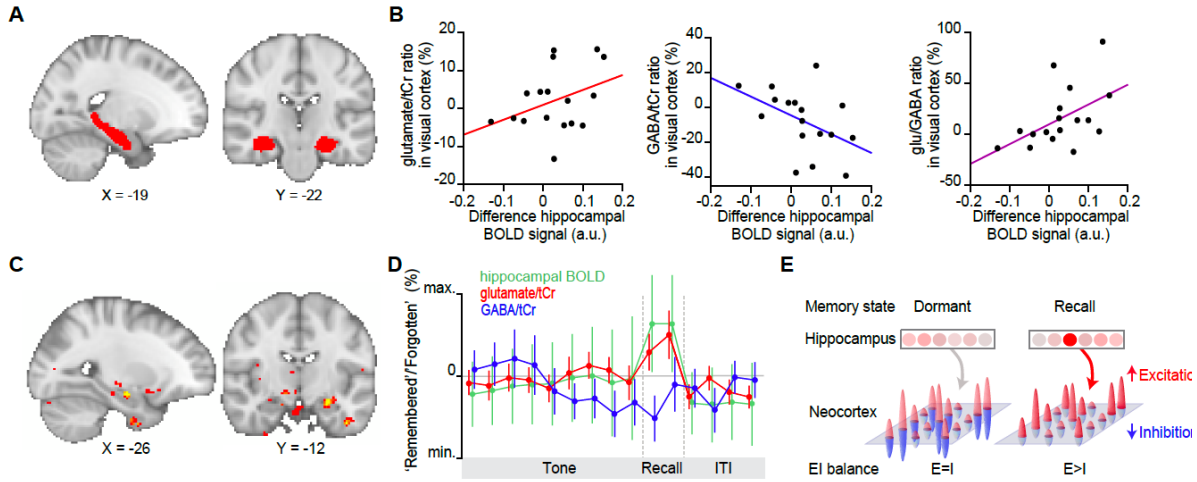


Figure 6.5 Hippocampal BOLD predicts neocortical glu/GABA ratio during recall
A Region of interest (ROI) in the hippocampus (red). MNI coordinates. **B** Across participants, the increase in hippocampal BOLD signal during 'remembered' compared to 'forgotten' trials positively predicted the decrease in GABA and the increase in glu/GABA ratio observed in V1 (Figure 6.4) (glutamate:tCr: r_{15}=0.14, p=0.585; GABA:tCr: r_{15}=-0.56, p=0.022; glu/GABA ratio: r_{15}=0.52, p=0.033). **C** Across the imaged brain volume, for 'remembered' versus 'forgotten' trials the correlation between BOLD signal and V1 glu/GABA ratio was selectively observed in the left hippocampus (t_{16}=11.37, p=0.005, whole-brain FWE corrected; Supplementary Table 6.7; MNI coordinates). **D** Moving average showing the ratio of 'remembered' to 'forgotten' trials during the inference test: hippocampal BOLD signal (green, range [-4:4]), glutamate:tCr (red, range [-8:8]), GABA:tCr (blue, range [-15:15]). Each point represents a 2.5s time bin (mean ± SEM). By incorporating a random jitter in the behavioural paradigm, MRS spectra across all trials and all participants were acquired in all possible 2.5 s time bins of the inference test trial (Supplementary Figure 6.3), thus achieving a higher temporal resolution than the TR of 4 s (see *Methods*). **E** Schematic illustrating how the hippocampus may facilitate (if indirectly) memory recall of a sensory cue during a transient break in neocortical E/I balance.

6.5 Discussion

The hippocampus is thought to provide an index for memories stored across distributed neocortical circuits (Goode et al., 2020; Teyler and DiScenna, 1985; Teyler and Rudy, 2007). However, the mechanism by which hippocampal activity coordinates with neocortex to facilitate memory recall has remained unclear. Here, using time-resolved fMRI-fMRS in humans, we show that recall of a visual cue is accompanied by a dynamic increase in the ratio between glutamate and GABA in visual cortex. This transient increase in glu/GABA ratio in visual cortex is selectively predicted by activity in the hippocampus. Accordingly, we propose the hippocampus gates recall of memories stored across distributed neocortical circuits using a disinhibitory mechanism (Figure 6.5E). This mechanism may explain how a memory index represented by the hippocampus selectively releases otherwise dormant representations stored across distributed neocortical circuits.

By simultaneously acquiring both fMRI and fMRS data, we provide a macroscopic readout of memory recall that reflects the consequence of underlying neural circuit level processes. Insight into the nature of these underlying circuit level processes can be gained from related data from

animal models. For example, the neural circuit mechanisms that underlie an increase in glu/GABA ratio during recall may be informed by evidence that the ratio between excitatory and inhibitory synaptic conductances in cortical neurons fluctuate around a stable set point (Anderson et al., 2000; Okun and Lampl, 2008; Wehr and Zador, 2003; Wilent and Contreras, 2005). This overall E/I proportionality ensures that neurons and networks are neither hypo- nor hyper-excitable for prolonged periods, allowing memories to be held in a dormant state (Barron et al., 2016a; Froemke et al., 2007; Vallentin et al., 2016; Vogels et al., 2011) that is protected from interference caused by new learning (Koolschijn et al., 2019; Kuchibhotla et al., 2017). However, despite overall proportionality, the exact E/I ratio is highly dynamic and transient breaks in E/I balance appear necessary for new learning and memory expression (Letzkus et al., 2015). Here, the reported fluctuations in MRS-derived glu/GABA ratio during memory recall may therefore reflect, if indirectly, dynamic changes in E/I balance.

Similarly, the reported relationship between the fluctuations in glu/GABA ratio and hippocampal activity may be informed by data from animal models. Of particular relevance are studies in rodents which show that glutamatergic projections from higher-order or interconnected brain regions can target disinhibitory cortical circuits to provide selective E/I modulation (Krabbe et al., 2019; Lee et al., 2013; Zhang et al., 2014). For example, to enhance visual discrimination during attentional modulation, projections from the cingulate region of mouse frontal cortex modulate activity in V1 by targeting vasoactive intestinal polypeptide-expressing (VIP+) interneurons, which in turn preferentially target other interneuron subtypes to release excitatory principle cells from inhibitory control (Zhang et al., 2014). During memory recall, hippocampal projections may similarly permit memory reinstatement by targeting disinhibitory circuits in neocortex. The correlation between hippocampal activity and glu/GABA ratio reported here may therefore reflect a mechanism whereby activity in the hippocampus facilitates cortical disinhibition to release otherwise latent cortical associations from inhibitory control.

This interpretation of the data is consistent with the notion that the hippocampus provides a memory index to flexibly reinstate information in extrahippocampal circuits (Goode et al., 2020; Teyler and DiScenna, 1985; Teyler and Rudy, 2007). Moreover, our findings replicate equivalent analyses conducted on fMRI data acquired using the same task (Barron et al., 2020) and are consistent with previous studies in humans showing evidence for coordinated hippocampal-neocortical memory reinstatement (Horner et al., 2015; Pacheco Estefan et al., 2019). When combined with the fMRS data, our results also corroborate findings in humans showing that hippocampal glutamate and GABA can predict mnemonic control (Nikolova et al., 2017; Schmitz et al., 2017). Taken together, we propose a mechanism for hippocampal indexing whereby

hippocampal projections control the release of mnemonic representations in sensory cortices by targeting disinhibitory circuits.

Given this interpretation of the data, we emphasise that the relationship between MRS-derived measures of glutamate and GABA and physiological measures of E/I balance remains complex. Rapid changes in synaptic glutamate and GABA that accompany neurotransmitter release occur on a timescale that is not possible to detect using fMRS. Moreover, only a fraction of MRS-derived neurometabolite concentration reflects neurotransmitter release. Of the different pools of glutamate and GABA (cytoplasmic, vesicular, or extracellular), MRS is considered most sensitive to unconstrained, intracellular metabolic pools that reside at relatively high concentration in the neuronal cytoplasm (Rae, 2014). Indeed, changes in extracellular GABA of less than 100-fold are unlikely to be detectable using MRS (Myers et al., 2016) and post-mortem studies suggest MRS is not sensitive to intracellular pools that reside in the mitochondria or vesicles (De Graaf and Bovée, 1990; Kauppinen and Williams, 1991).

Interpretation of MRS-derived glutamate and GABA is further complicated by the fact that the release and recycling of glutamate and GABA constitute major metabolic pathways (Bak et al., 2006; Magistretti and Allaman, 2015). Yet, the metabolic and neurotransmitter pools are thought to be tightly coupled during anaesthesia, rest and certain stimulation protocols, with a 1:1 relationship reported between the rate of glutamine-glutamate cycling, which is necessary for glutamate and GABA synthesis, and neuronal oxidative glucose consumption, which indirectly supports neurotransmitter release among other processes (Rothman et al., 2003; Shen et al., 1999; Sibson et al., 1998). Therefore, an increase in synaptic neurotransmission occurs together with an increase in synthesis of exogenous glutamate, which provides a precursor for GABA via the glutamate-glutamine cycle. During sensory stimulation a transient uncoupling has been observed with a short-lived mismatch between glucose utilisation and oxygen consumption (Fox et al., 1988; Fox and Raichle, 1986), particularly during stimulation protocols that alternate between high intensity and quiescent periods (Gjedde et al., 2002). Dynamic fluctuations in fMRS-derived glutamate and GABA reported here may therefore reflect transitions to new metabolic steady states (Stanley and Raz, 2018), which could reflect (if indirectly) relative shifts in E/I equilibrium at the physiological level. During associative memory recall, the increase in glu/GABA ratio may therefore be interpreted as an increase in synthesis of glutamate relative to degradation, with an opposing effect on GABA.

This interpretation is supported by a handful of previous studies showing event-related changes in MRS glutamate (Apšvalka et al., 2015; Gussew et al., 2010; Lally et al., 2014) and GABA (Cleve

et al., 2015), together with a growing body of evidence reporting a relationship between MRS-derived measures of neurometabolites and behaviour (Puts et al., 2011; Scholl et al., 2017; Stagg et al., 2011). Nevertheless, it remains to be established whether unconstrained glutamatergic and GABAergic pools show event-related changes that are MRS-sensitive. To validate this interpretation of event-related fMRS it is important to leverage animal studies where more sensitive methods can be employed to relate fMRS measures to physiological parameters. Here, by implementing an inference task in VR, we operationalise memory recall using the exact same paradigm previously employed in rodents (Barron et al., 2020). Therefore, in addition to engaging memory-dependent inference, "opening the box" to find a reward in the VR environment approximated the process of rodents finding a reward from a dispenser in a 3D environment. By using VR, the findings presented here may be compared to data acquired in animal models in ongoing future research. In this manner, VR paradigms in humans may provide a basis from which to gain insight into the cellular and circuit mechanisms that underlie macroscopic measures of excitation and inhibition. This may prove particularly useful for establishing a more detailed understanding of the relationship between fMRS-derived measures of glutamate and GABA and physiological measures of E/I balance.

Previous MRS protocols typically employ a 'block' design, where a static measure of the concentration of glutamate and GABA is achieved by averaging the spectra across a time-window that may span several minutes. This approach obscures the temporal dynamics of neurometabolites which more closely relate to fluctuations in excitation and inhibition reported at the physiological level. Similarly, dynamic changes in neurometabolites that accompany cognitive processes and ongoing behaviour are overlooked. Indeed, when the average concentration of Glx and GABA are considered in V1 across time, no significant relationship is observed across subjects (Rideaux, 2021), a result which we also observed when assessing average glutamate and GABA using our dataset. By contrast, with the increase in availability of ultra-high field MRI scanners and the development of more advanced sequences (Stagg and Rothman, 2013), fMRS has emerged as a viable method to detect dynamic changes in neurochemicals in both healthy and clinical populations (Stanley and Raz, 2018).

Although there are currently only a handful of event-related fMRS studies, together with our data, these suggest fMRS is highly sensitive to detecting task-relevant dynamic changes in glutamate and GABA (Jelen et al., 2018). For example, in the lateral occipital complex fMRS demonstrates differences in glutamate in response to presentation of objects versus abstract stimuli (Lally et al., 2014), and in the left anterior insula fMRS reveals a transient increase in glutamate with exposure to painful stimuli (Gussew et al., 2010). fMRS-derived glutamate is even sufficiently

sensitive to detect repetition suppression effects in the lateral occipital complex (Apšvalka et al., 2015), mirroring analogous effects reported in fMRI (Barron et al., 2016a; Grill-Spector et al., 2006). Here, we further illustrate that within a 3 second time window delineated by the question period in the inference task, the temporal resolution of fMRS is sufficient to relate transient changes in glutamate and GABA to associative memory recall. Importantly, we compare data across two conditions ('remembered' and 'forgotten') to inherently control for: *(1)* between-subject differences in average GABA and glutamate which are affected by demographic (e.g. age and gender); *(2)* between-subject differences in spectral quality; *(3)* between-subject differences in tissue composition; *(4)* between-subject differences in the effect of other neurochemicals on measures of glutamate and GABA. Such time-resolved, within-subject and condition-dependent fMRS may provide a promising tool to capture real-time, task-relevant changes in neurometabolites, on a timescale equivalent to task-based fMRI. Assessing whether the temporal resolution of fMRS can be further improved will likely prove an important step in refining fMRS in the future.

During associative memory recall, the transient increase in glu/GABA ratio reported in our data can be accounted for by a significant decrease in the concentration of MRS-derived GABA. Notably, detecting dynamic changes in GABA is challenging for two key reasons: the concentration of GABA in human brain tissue is relatively low and the spectral peaks for GABA overlap with other, more abundant neurochemicals (Andreychenko et al., 2012; Govindaraju et al., 2000; Puts and Edden, 2012). While the most common approach to detecting MRS-derived GABA involves using a J-difference spectral editing technique to separate GABA peaks from overlapping peaks (Bottomley, 1987; Mescher et al., 1998), here we use a non-edited sequence (sLASER). While spectral editing may provide higher precision (Hong et al., 2019), this occurs at the cost of a larger volume of interest, longer TEs and higher susceptibility to motion and drift artefacts due to longer acquisition times, making it less suitable for event-related fMRS (Terpstra et al., 2006; Trabesinger and Boesiger, 2001). Moreover, direct comparisons between edited and non-edited sequences at 7T reveal no significant difference in the concentration of GABA measurements (Hong et al., 2019). Therefore, together with studies reporting dynamic changes in GABA with sensory stimulation (Lin et al., 2012; Mekle et al., 2017), our data illustrates how a non-edited sequence can provide sufficient data quality for measuring dynamic changes in MRS-derived GABA, which cannot be explained by changes in compounds at higher-concentration that have overlapping peaks (i.e. glutamate or NAA, Supplementary Figure 6.9). Indeed, Monte Carlo simulations reported here validate that non-edited sequences can be used to quantify dynamic changes in GABA (Figure 6.4G; Supplementary Figure 6.4). Moreover, compared to spectral editing, our

approach comes with the advantage of simultaneously measuring dynamic changes in GABA and glutamate, together with 17 other neurometabolites.

To detect dynamic changes in GABA, it was necessary to disable default priors on the spectral fitting procedure that constrain GABA as a ratio to more stable metabolite concentrations (Supplementary Figure 6.4). As a consequence, we were able to detect dynamic changes in both glutamate and GABA across time, as illustrated using Monte-Carlo simulations and permutation testing. By comparing the change in metabolite concentration between two conditions ('remembered' versus 'forgotten'), the ratio in GABA between conditions rather than absolute values was the key measure of interest. However, we note that absolute GABA estimates were higher compared to those obtained using default priors that normalise estimates relative to more stable metabolite concentrations. Importantly, the quality of our MRS data was comparable with other studies that have acquired 7T MRS data from visual cortex (Bednařík et al., 2018; Hong et al., 2019; Mekle et al., 2017; Prinsen et al., 2017). Moreover, the quality of the glutamate estimates was in line with previous studies employing event-related fMRS to assess dynamic changes in glutamate (Apšvalka et al., 2015; Gussew et al., 2010; Lally et al., 2014).

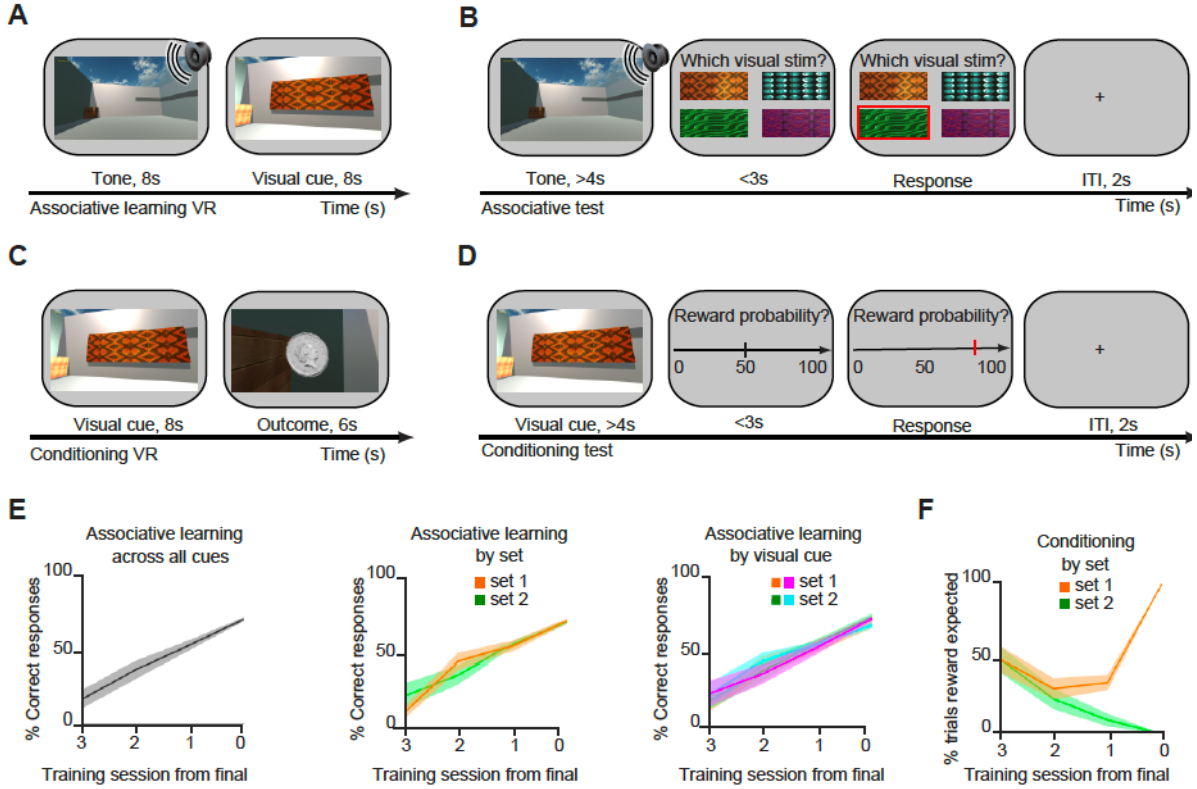
Disturbances in E/I balance are thought to underlie a number of neuropsychiatric conditions, including schizophrenia, autism, epilepsy and Tourette's syndrome (Robertson et al., 2016; Stanley and Raz, 2018; Taylor et al., 2015). While previous studies report inconsistencies in MRS-derived measures of glutamate and GABA in these patient populations, this may be attributed to differences in brain region, cognitive state and imaging protocol, among other factors. Here, by using both fMRS and fMRI to reveal a signature change in glu/GABA ratio that relates to hippocampal BOLD signal, behavioural performance and cognition, our findings present a potential target for clinical investigation. Moreover, our findings show that even in the healthy brain a transient break in E/I balance is necessary to support key cognitive processes such as memory recall.

In summary, using time-resolved fMRI-fMRS we report a transient increase in glu/GABA ratio in V1 during associative recall of a visual cue. This increase in neocortical glu/GABA ratio is predicted by activity in the hippocampus. By unveiling this coordination between the hippocampus and neocortex, we show how the hippocampus may have the capacity to selectively modulate and disinhibit memories represented in neocortex. This mechanism may explain how the hippocampus plays a key role in memory recall, by indexing the release of specific memories stored across distributed neocortical circuits.

6.6 Conclusion

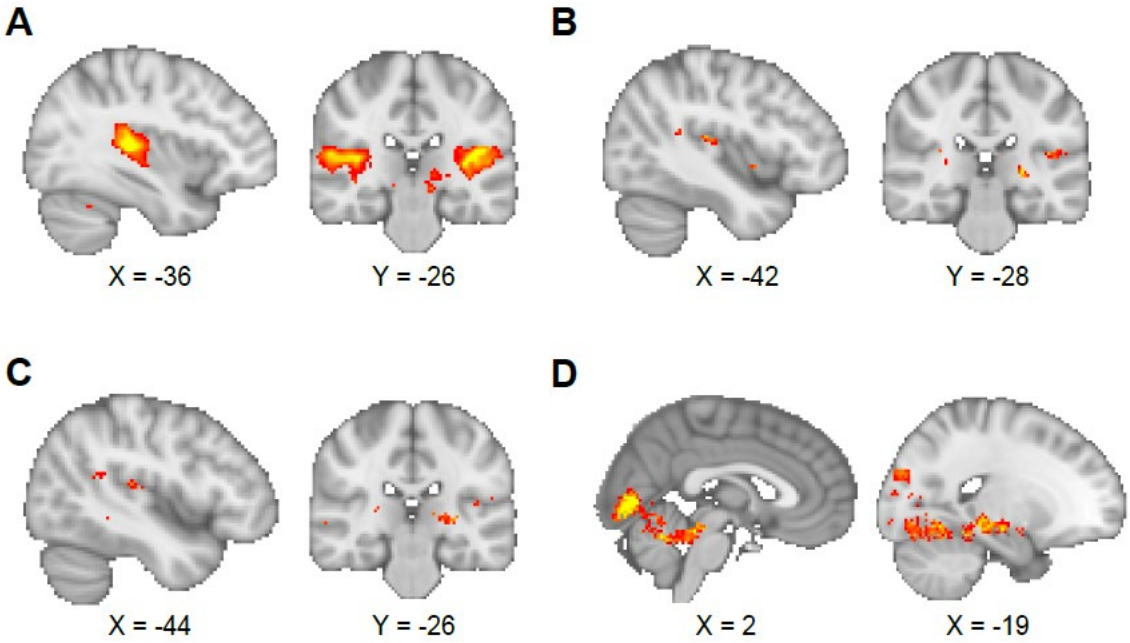
In this chapter I used a novel MRI sequence to simultaneously acquire fMRI and time-resolved fMRS during a memory recall task. I showed that memory recall is associated with a dynamic decrease in cortical GABA, and that this decrease in inhibition is predicted by activity in hippocampus. These results reveal how hippocampus and neocortex interact to release memories from inhibition.

6.7 Supplementary figures

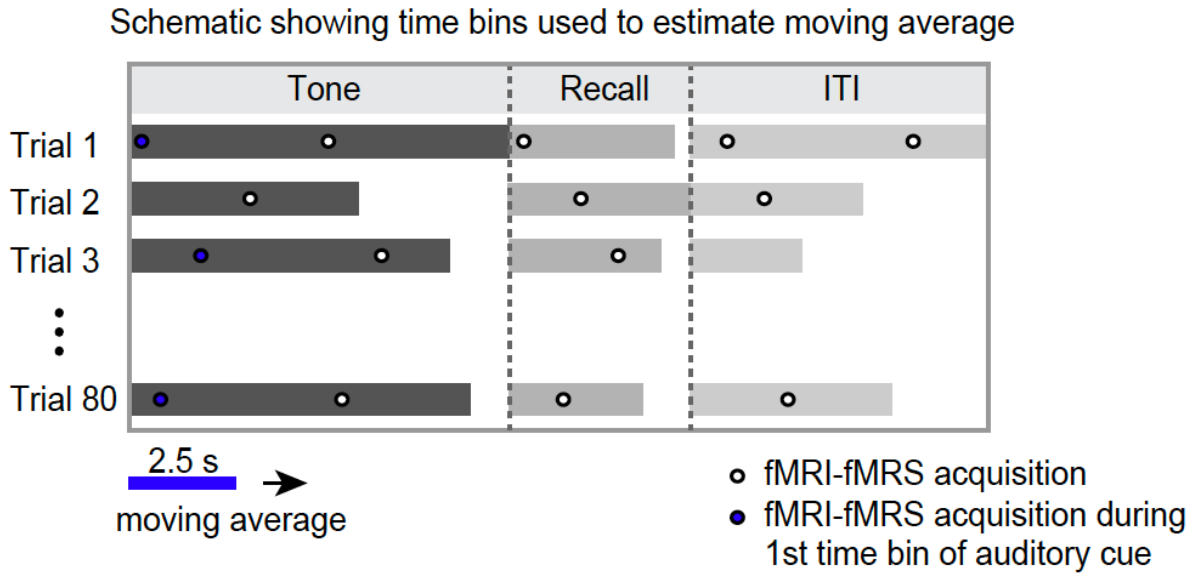


Supplementary Figure 6.1 Behavioural training and performance

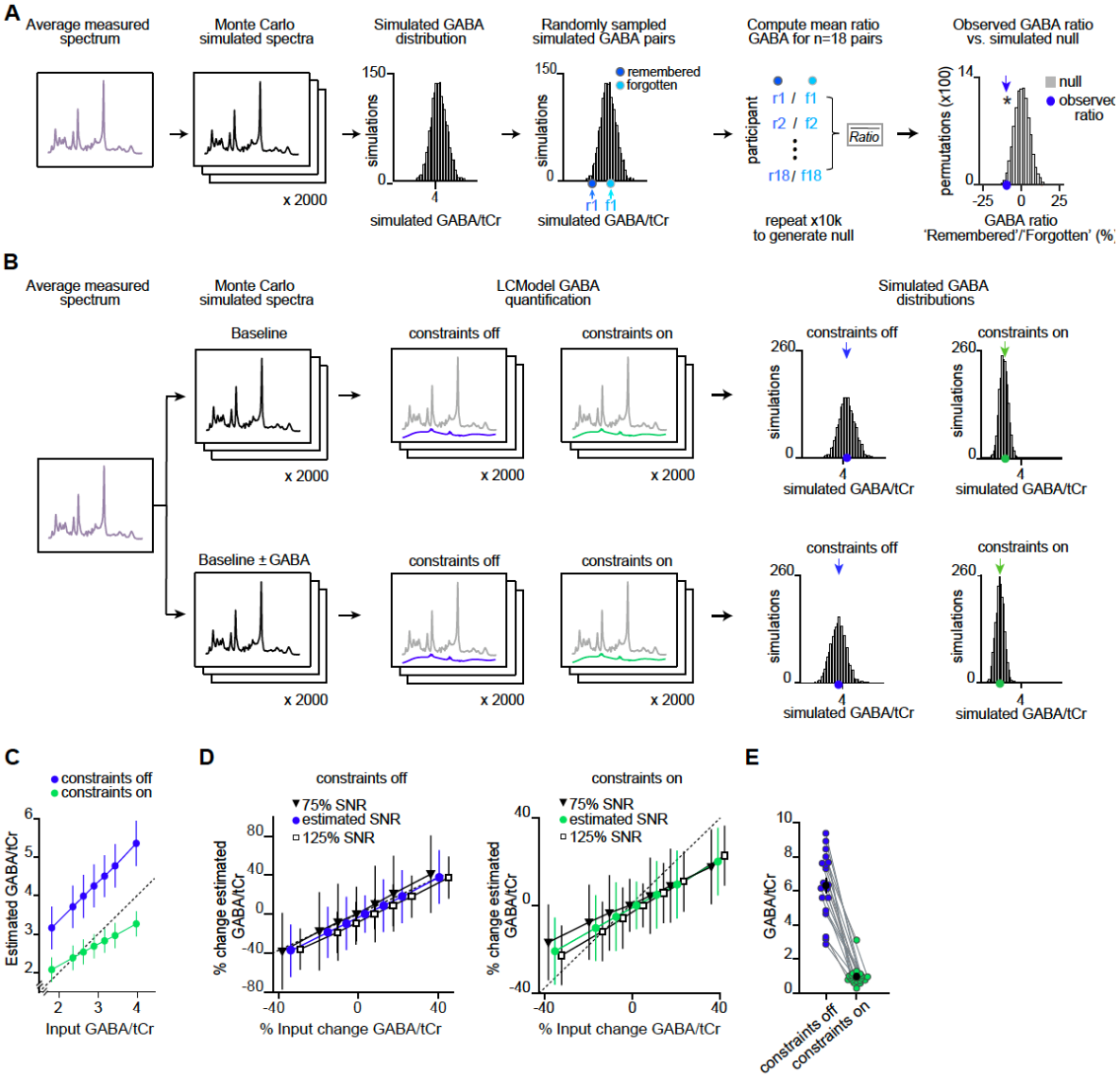
A-B On day 1, during the 'associative learning' stage, participants learned to associate each of the 80 different auditory cues with one of 4 possible visual cues. **A** The associative learning stage was performed in a VR environment (Figure 6.1A). **B** Learning was monitored using an associative test, in the absence of feedback. **C-D** On day 2, during the 'conditioning' stage, participants learned to associate each of the four visual cues with one of two possible outcomes (monetary reward for set 1; neutral woodchip for set 2) (Figure 6.1B). **C** The conditioning stage was performed in a VR environment (Figure 6.1A). **D** Learning was monitored using a conditioning test, in the absence of feedback. **E** Left: Participants performed the associative learning task until they showed recall accuracy on the associative test of at least 50% of the 80 possible auditory-visual pairs. Middle: On the associative test, used to monitor performance during the associative learning, there was no difference in accuracy between auditory-visual pairs in set 1 (rewarded) and set 2 (neutral). Right: On the associative test there was no difference in accuracy between auditory cues associated with the four different visual cues. **F** On day 2, participants performed the conditioning task until they reached 100% accuracy on all visual-outcome associations in the conditioning test.



Supplementary Figure 6.2 Comparison of different smoothing parameters applied to fMRI data
A-C The BOLD signal response to auditory cues in the inference test was used to assess the effect of different smoothing parameters (contrast of interest: [all trials during inference test – all trials during conditioning]). As noted in the *Methods*, the quality of the fMRI data in the interleaved fMRI-fMRS sequence was compromised relative to contemporary standards for 7T fMRI. The smoothing parameters applied to the data influenced the reliability of the analysis, as illustrated here. MNI coordinates. **A** Smoothing at the second-level using a 5 mm kernel, approximately two times the voxel size, as recommended (Worsley and Friston, 1995), gave significant BOLD signal in bilateral auditory cortex (left auditory cortex, t_{17}=8.94, p<0.001; right auditory cortex, t_{17}=6.76, p<0.001; whole-brain FWE corrected; thresholded at p<0.001 uncorrected for visualisation purposes only). This smoothing protocol was used for all analyses presented in the main figures. **B** Smoothing at the first-level using a 5 mm kernel, did not give significant BOLD signal in auditory cortex (whole-brain FWE corrected; thresholded at p<0.01 uncorrected for visualisation purposes only). **C** Smoothing at the first-level using an 8 mm kernel (default for SPM) did not give significant BOLD signal in auditory cortex (whole-brain FWE corrected; thresholded at p<0.01 uncorrected for visualisation purposes only). **D** When contrasting 'remembered' and 'forgotten' trials in the inference test, smoothing at the first-level using an 8 mm kernel (as in *C*) gave similar results to those presented in Figure 6.3C, despite the absence of a main effect response to auditory cues as shown in *C*. Thus, as shown in Figure 6.3C, during the question period in the inference test trials (Figure 6.1B,D), BOLD signal in the visual cortex and the hippocampus was significantly higher for 'remembered' versus 'forgotten' auditory cues ('remembered' – 'forgotten', visual cortex: t_{17}=6.01, p<0.001; right hippocampus: t_{17}=7.58, p<0.001; left hippocampus (sub-threshold): t_{17}=4.91, p>0.05; whole-volume FWE-corrected; thresholded at p<0.01 uncorrected for visualisation purposes only; MNI coordinates).



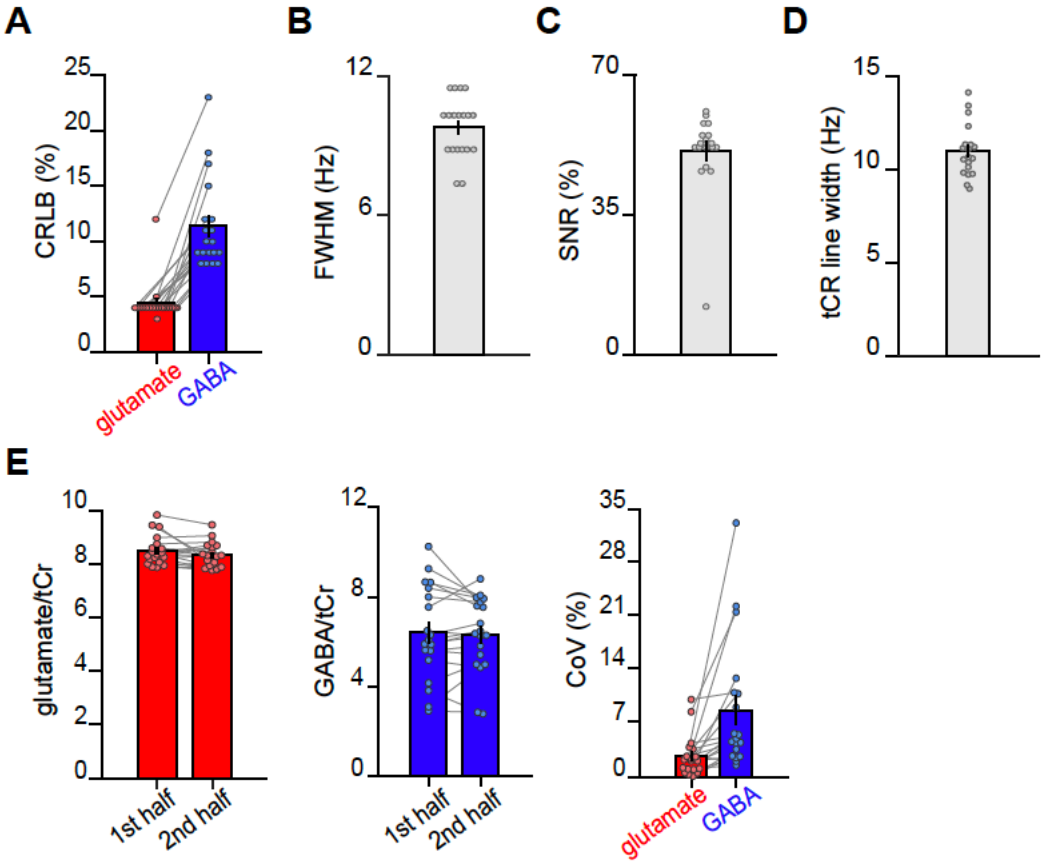
Supplementary Figure 6.3 Estimation of fMRS moving average
Illustration of how the moving average for fMRS data was generated. To generate the moving average for fMRS data in Figure 6.4, 6.5 and Supplementary Figure 6.11, we leveraged the jitter in the temporal relationship between data acquisition and experimental paradigm. Data acquisition and experimental paradigm were started at the same time. Due to the temporal jitter in each trial of the experimental paradigm (see *Methods*) fMRS spectra were acquired across different segments of the inference test trials. Each time bin constituted a 2.5 s time window shifted by 0.5 s in each iteration. For each subject, spectra across all 80 trials were averaged within 2.5 s time bins to generate a moving average. The first time bin during the auditory cue ('Tone') is shown as an example in blue.



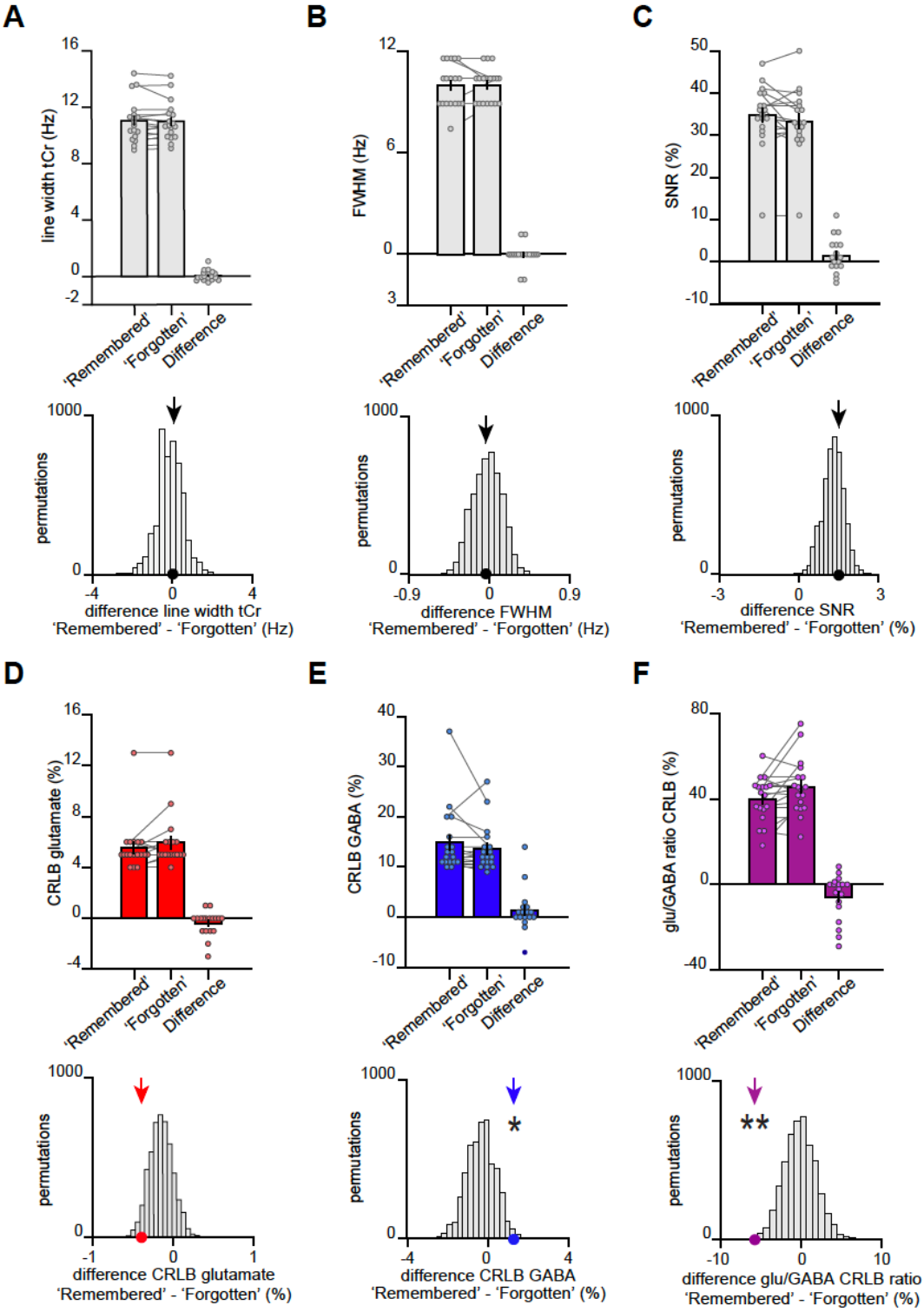
Supplementary Figure 6.4 Monte Carlo simulations to assess MRS data quality

A-B The average measured spectra were used as an input to Monte Carlo simulations, to generate simulated spectrum with the level of noise (SNR) and linewidth matched to the observed data. **A** Using this simulated data we established a null distribution for the GABA:tCr ratio between 'remembered' and 'forgotten' spectra that would be expected by chance. Relative to this null distribution, the observed GABA ratio was significant (p=0.019), as shown in Figure 6.4G. **B** Using the Monte Carlo simulations, sets of 2000 spectra were generated with changes in GABA imposed relative to the average estimated GABA concentration ('Baseline'). The imposed changes in GABA included multiples (±0.5, ±1, ±2) of the observed difference between 'remembered' and 'forgotten' conditions shown in Figure 6.4E. The imposed change when subtracting 1x the observed difference between 'remembered' and 'forgotten' conditions is illustrated here in the lower row. All simulated spectra were then fit using LCModel to obtain a distribution of estimated GABA:tCr concentrations. We assessed how sensitive different model fitting settings in LCModel are to detecting a change in GABA. Specifically, we changed the NRATIO parameter to compare settings where GABA:tCr estimates are constrained relative to the concentration of more abundant neurochemicals ('constraints on', NRATIO=12), to settings where GABA estimates are not subject to prior constraints as used for analyses reported in Figs. 4-5 ('constraints off', NRATIO=0) (see Appendix 1–supplementary note). **C** For different imposed changes in GABA, the mean ± standard deviation for the resulting GABA:tCr concentrations (n = 2000) is shown, as estimated with LCModel using 'constraints off' (blue) and 'constraints on' (green). The dashed line represents agreement between the imposed and estimated GABA:tCr concentration. As expected, the estimated GABA:tCr was higher for 'constraints off' compared to 'constraints on'. Importantly, for 'constraints off' compared to 'constraints on', the range in estimated GABA:tCr concentrations more closely reflected the imposed change in GABA ('constraints off': slope=1.01; 'constraints on': slope=0.546). To estimate the power to reject the null hypothesis of equal slopes between conditions ('constraints off' versus 'constraints on', we randomly selected n=18 spectra for each imposed change in GABA per condition, thus

matching the number of simulations to the number of participants in our sample (Figure 6.4). Over 500 sets of size n=126 (i.e. n=18 simulated spectra for each imposed change in GABA), we estimated the power to reject the null hypothesis of equal slopes to be 0.997. Similarly, over 500 sets of size n=126 we estimated the t-statistic to reject the null hypothesis of equal slopes to be (t_{248}=5.23, p<0.001). These Monte Carlo simulations demonstrate that sensitivity to detecting dynamic changes in GABA:tCr is greater when using 'constraints off' compared to 'constraints on' in LCModel. **D** To verify whether changes in GABA could be detected in fMRS data with lower or higher average SNR, we looked at the effect of decreasing and increasing SNR in the simulated spectra upon detecting the imposed changes in GABA concentration using LCModel. With 'constraints off', the estimated changes closely reflected the imposed changes in GABA at different SNRs (slope=1.06 for 75% SNR; slope=0.985 for 125% SNR; power to reject null hypothesis of equal slopes estimated as described in *c*, '75% SNR' vs '125% SNR': 0.251, see *Methods*). With 'constraint on' the measured changes deviated from the imposed changes at both higher and lower SNRs (slope=0.458 for 75% SNR; slope=0.612 for 125% SNR; power to reject the null hypothesis of equal slopes '75% SNR' vs '125% SNR': 0.827, see *Methods*). These simulations show that dynamic changes in GABA are reflected in the measured GABA concentrations using 'constraints off', even when SNR is reduced relative to that observed in the data reported in Figure 6.4-5. **E** Comparison of GABA:tCr estimates with and without model constraints. When using a model fit that does not constrain GABA estimates within a predefined range ('constraints off'), the absolute concentration of GABA:tCr is higher and the resulting dynamic range is greater. When instead using a model fit that constrains metabolite values within a predefined range ('constraints on'), GABA:tCr values are biased towards the predetermined ratio and show reduced variance. * indicates p<0.05 *** indicates p<0.001.



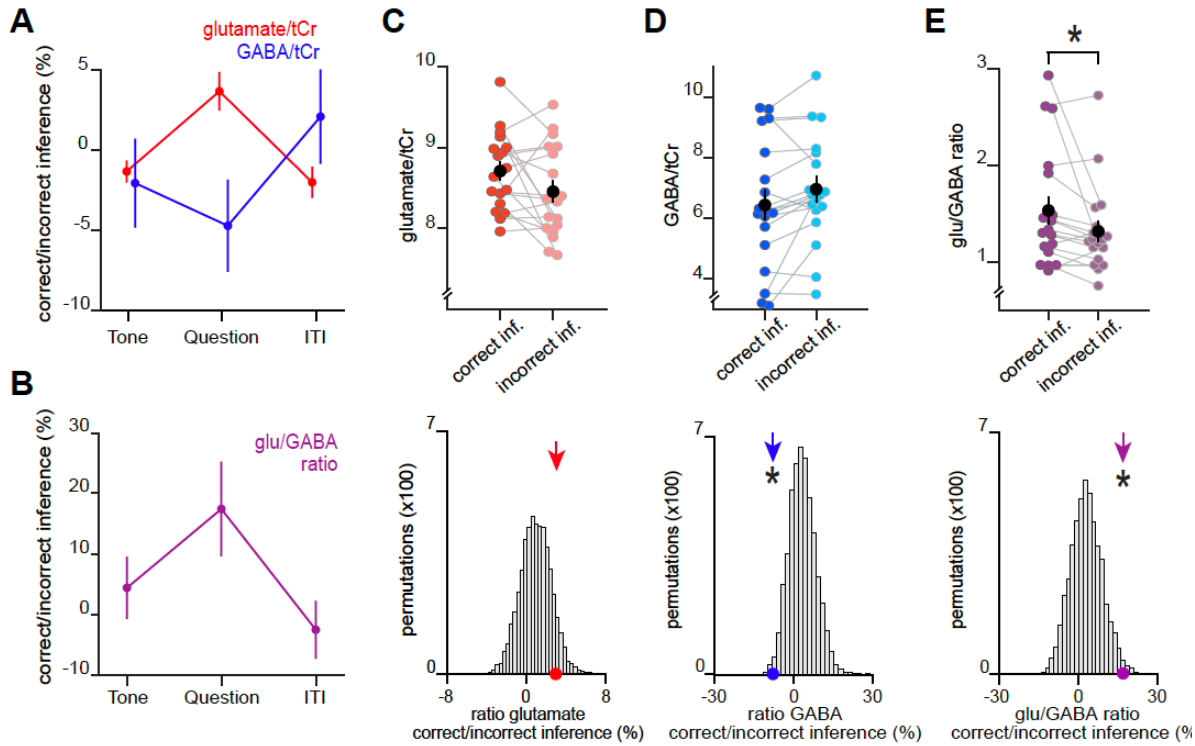
Supplementary Figure 6.5 MRS data quality metrics across all spectra.
For each subject, quality metrics across all MRS spectra were assessed (mean total number of spectra: 457). **A** The average Cramér-Rao Lower Bound (CRLB) for glutamate was 4.42±0.43% and the average CRLB for GABA was 11.37±0.94%. (mean ± SEM) **B** The average full width at half max (FWHM) as determined by LCModel was 9.86±0.30 Hz (mean ± SEM). **C** The average Signal to Noise Ratio (SNR) as determined by LCModel was 51.1± 2.37%. (mean ± SEM). **D** The average line width of the total Creatine (tCr) peak was estimated to be 11.01±0.32Hz. (mean ± SEM). **E** The intra-subject coefficient of variance (CoV) was estimated by splitting the dataset into two equal halves (on average 278 spectra in each half) and analysing each half in LCModel. We defined CoV as the standard deviation between the 2 halves divided by their mean. The intra-subject CoV for glutamate was 2.68±0.62%, and the intra-subject CoV for GABA was 8.74±1.95% (mean ± SEM). These findings demonstrate stability in our MRS measurements over the course of the scan task. Notably, this analysis differs from standard estimates of intra-subject CoV where test-retest is assessed across two separate scanning sessions.



Supplementary Figure 6.6 The transient break in glu/GABA ratio observed during recall cannot be explained by changes in data quality metrics or goodness of model fit

A-B At 7T, increases in BOLD effects alter T2* of metabolite signals, a phenomenon that results in line narrowing of all signals in the spectrum (Mangia et al., 2006). This phenomenon is most discernible on the strongest singlets such as total creatine (tCr). To assess the reliability of our fMRS measures we therefore quantified the difference in line width between our conditions of interest. These analyses show no significant difference in line width between categories 'remembered' and 'forgotten'. Thus, as line width was matched across categories there was no evidence for a category-specific bias in BOLD-related confounds for metabolite values reported in Figure 6.4. **A** The line width of the tCr peak was estimated for each participant for both the 'remembered' and 'forgotten' spectra acquired during the question period of inference test. Across participants we observed no difference in tCr line width between our two conditions of interest ('remembered' – 'forgotten': t_{17}=0.61, p=0.552) (mean ± SEM). To control

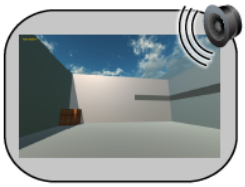
for systematic differences in the number of 'remembered' and 'forgotten' trials we compared the group mean difference in tCr (black arrow) against a null distribution generated by permuting the trial labels (x5000) and re-estimating line width difference for 'remembered' - 'forgotten'. Again, there was no significant difference between the group mean difference in tCr and the null distribution (p=0.403). **B** To further quantify differences in line width for each participant's 'remembered' and 'forgotten' spectra, we compared the full width at half maximum (FWHM). Between our two conditions of interest, we observed no significant difference in FWHM ('remembered' – 'forgotten': t_{17}=-0.21, p=0.832) (mean ± SEM). There was no significant difference between the FWHM group mean for 'remembered' – 'forgotten' and its null distribution generated as described in (*a*) (p=0.466). **C** To verify that the differences between 'remembered' and 'forgotten' cannot be attributed to a difference in signal strength, we assessed the SNR for both conditions. We observed no significant difference in SNR ('remembered' - 'forgotten': t_{17}=1.55, p=0.140). In addition, there was no significant difference between the SNR group mean difference for 'remembered' – 'forgotten' and its null distribution (p=0.319). The positive shift of the null distribution can be explained by the difference in number of trials between the 'remembered' and 'forgotten' conditions (Supplementary Table 6.6). **D-F** CRLB estimates can be applied to confirm the effects we observed in Figure 6.4D-F, i.e. that glutamate:tCr and glu/GABA ratio increase and GABA:tCr decreases for 'remembered' vs 'forgotten' spectra. CRLB % estimates decrease (indicating more confident model fit) as metabolite concentrations increase, and vice versa. For the difference between conditions ('remembered' – 'forgotten'), the increase in the relative concentration of glutamate has a negative effect on CRLBs, while the decrease in the relative concentration of GABA have a positive effect on CRLBs. Indeed, we observed a decrease in glu/GABA CRLB, supporting the observed increase in glu/GABA ratio for 'remembered' vs 'forgotten' ('remembered' - 'forgotten', glutamate:tCr: t_{17}=-1.69, p=0.110; GABA:tCr: t_{17}=1.28, p=0.218; glu/GABA ratio: t_{17}=-2.29, p=0.035). Next, we compared the true group mean CRLBs ('remembered' – 'forgotten') with null distributions generated by permuting the trial labels. Relative to these respective null distributions, we asked whether the true difference in CRLBs for 'remembered' 'forgotten' was negative for glutamate and glu/GABA ratio, and positive for GABA. For glutamate and glu/GABA ratio we observed significant differences between the CRLB group means and their respective null distributions in the expected direction: CRLBs for Glutamate and glu/GABA ratio were significantly negative (glutamate:tCr: p=0.071; GABA:tCr: p=0.011; glu/GABA ratio: p=0.002). These control analyses support the differences between 'remembered' and 'forgotten' spectra we report in Figure 6.4D-F. * indicates p<0.05, ** indicates p<0.01.



Supplementary Figure 6.7 An increase in glu/GABA ratio in V1 during memory recall are also observed when categorising trials into 'remembered' and 'forgotten' using a less conservative approach

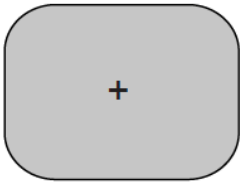
Here we repeated the analyses in Figure 6.4 using only performance on the inference test to categorise trials from the inference test into 'remembered' (correct inference) and 'forgotten' (incorrect inference). **A-B** Similar to using the more conservative definition for 'remembered' and 'forgotten' shown in Figure 6.4, glu/GABA ratio significantly increased during the question period for 'remembered' versus 'forgotten' trials ('correct inference':'incorrect inference', glu/GABA ratio: t_{17}=2.23, p=0.039). One participant (n=1) was excluded from this analysis, where there were less than 8 spectra for either the 'remembered' or 'forgotten' conditions during the question period. This break in glu/GABA ratio was not observed during the 'tone' (~7 s) or 'ITI' (~2.7 s) periods ('Tone', glu/GABA ratio: t_{18}=0.87, p=0.395; 'ITI', glu/GABA ratio: t_{18}=-0.51, p=0.616). **C-E** Upper row: Metabolite values and glu/GABA ratio during the question period for 'remembered' and 'forgotten' trials (mean ± SEM). Lower row: Comparison of the mean ratio of 'correct inference' to 'incorrect inference' (coloured arrows) against null distributions generated by permuting the trial labels **(**x5000**)** to control for any potential biases in the analyses. Relative to the null distributions, GABA significantly decreased while glu/GABA ratio significantly increased (glutamate:tCr: p=0.088; GABA:tCr: p=0.017; glu/GABA ratio: p=0.011). * indicates p<0.05.

During auditory cue



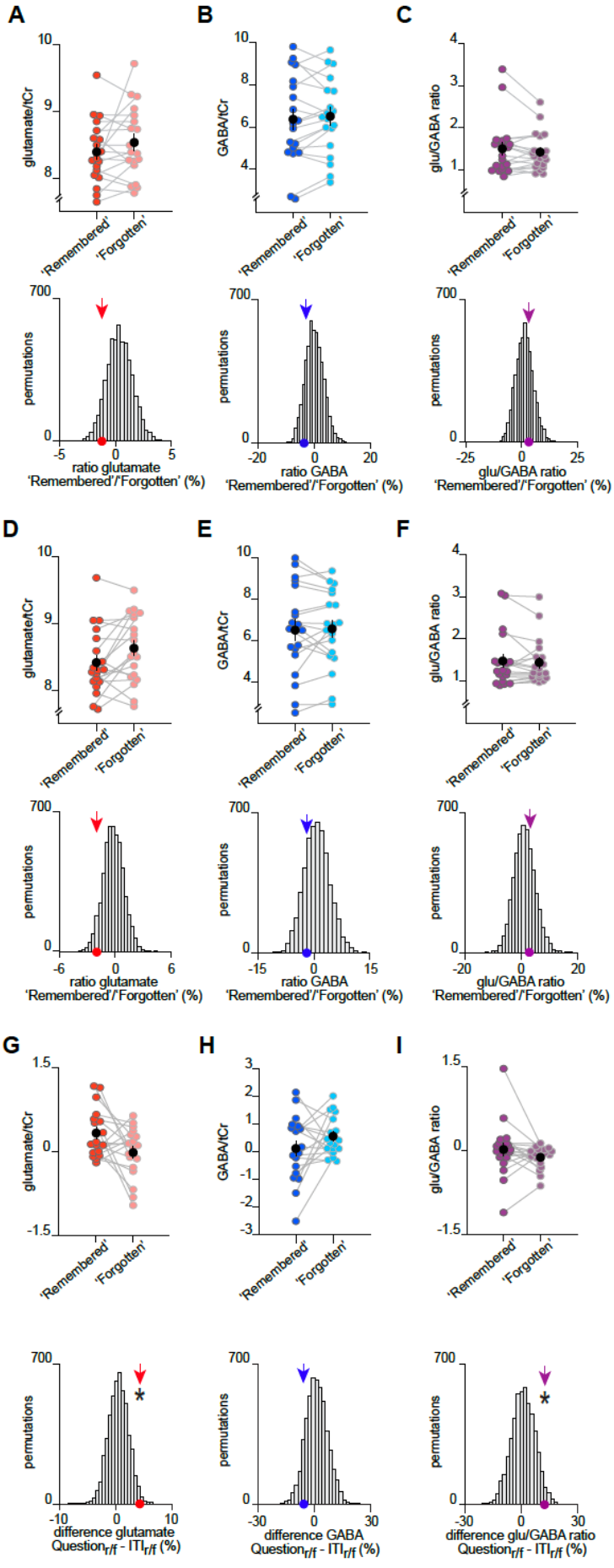
Tone, >4s

During the ITI



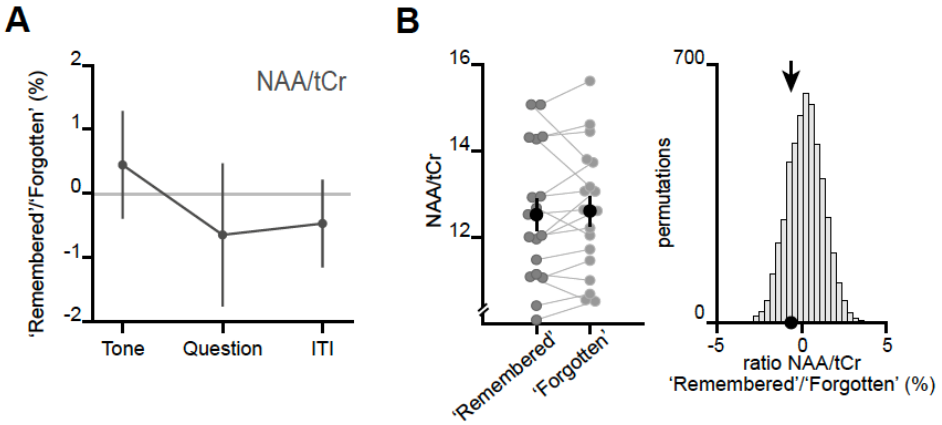
1.4-10s, mean 2.7s

Difference: Question - ITI



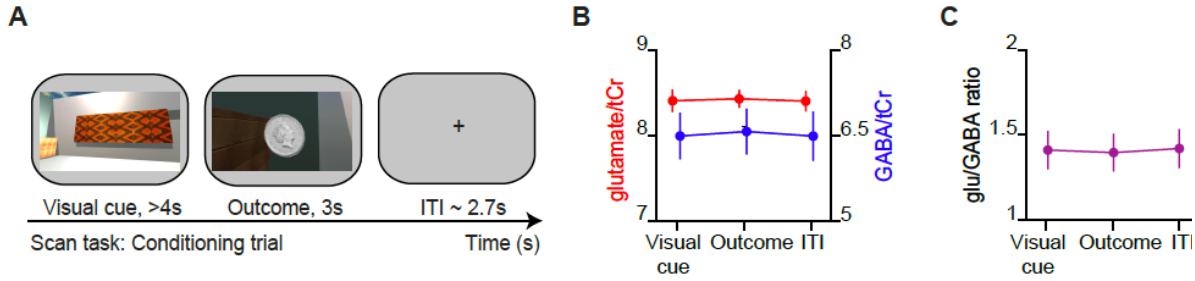
Supplementary Figure 6.8 The change in glu/GABA ratio is transient and only observed during memory recall

A-F Upper rows: The difference in metabolite ratios for 'remembered' versus 'forgotten' trials during the auditory cue period (a-c) and ITI period (d-f) (mean ± SEM). Lower rows: To control for any biases due to differences in the number of 'remembered' and 'forgotten' trials, we compared the group mean (coloured arrows) against a null distribution generated by permuting the trial labels (x5000). MRS voxel shown in Figure 6.3D. **A-C** During the auditory cue period, there was no significant increase in glu/GABA ratio between 'remembered' and 'forgotten' trials ('remembered':'forgotten', glutamate:tCr: t_{18}=-1.34, p=0.198; GABA:tCr: t_{18}=1.11, p=0.282; glu/GABA ratio: t_{18}=1.01 , p=0.325). Similarly, there was no significant difference between any of the group means and their respective null distributions (glutamate:tCr: p=0.107; GABA:tCr: p=0.191; glu/GABA ratio: p=0.314). **D-F** During the ITI period, there was no significant increase in glu/GABA ratio between 'remembered' and 'forgotten' trials ('remembered':'forgotten', glutamate:tCr: t_{18}=-2.27, p=0.036; GABA:tCr: t_{18}=-0.44, p=0.661; glu/GABA ratio: t_{18}=0.40, p=0.692). For GABA and glu/GABA ratio, there were no differences between any of the group means and their respective null distributions, although a significant decrease was observed for glutamate (glutamate:tCr: p=0.029; GABA:tCr: p=0.223; glu/GABA ratio: p=0.386). **G-I** To test whether the break in glu/GABA ratio was transient and only observed during memory recall, we compared our measure of glu/GABA ratio for 'remembered':'forgotten' during the question period ('Question') to the period immediately after ('ITI') on the same trial. To control for the difference in numbers of trials between conditions, we compared the difference of the group means to a permuted null distribution. Compared to the respective null distributions, a significant difference was observed for glutamate and glu/GABA ratio between the group means for 'Question' versus 'ITI' (glutamate:tCr: p=0.017; GABA:tCr: p=0.124; glu/GABA ratio: p=0.033). 'r/f' indicates 'remembered':'forgotten'; *indicates p<0.05.

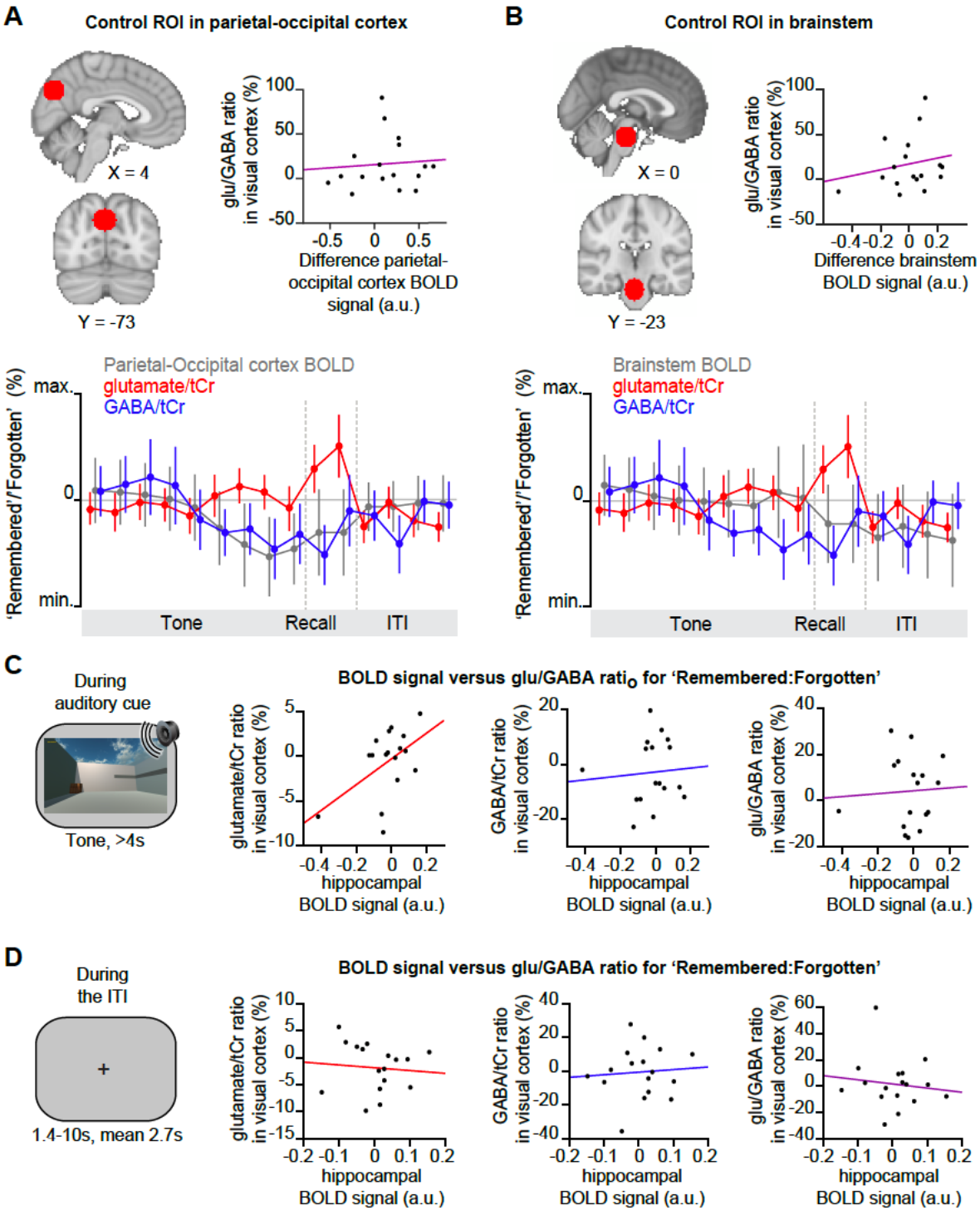


Supplementary Figure 6.9 The changes in metabolite concentrations cannot be attributed to changes in NAA:tCr

A No significant change between 'remembered' and 'forgotten' trials was observed for NAA during the 'Question' period of inference trials (remembered:forgotten NAA:tCr: t_{17}=-0.58, p=0.570). Notably, NAA has overlapping peaks with GABA but is found at higher concentration. **B** Left: Metabolite values during the question period for 'remembered' and 'forgotten' trials (mean ± SEM). Right: Ratios between 'remembered':'forgotten' (black arrows) against null distributions generated by permuting the trial labels. Relative to the null distributions, the mean NAA ratio showed no significant difference (NAA:tCr: p=0.220).



Supplementary Figure 6.10 During conditioning trials, no difference in glu/GABA ratio was observed
A Example conditioning trial encountered during the MRI scan task. **B** The concentration of glutamate:tCr and GABA:tCr during the 'visual cue', 'outcome' and 'ITI' periods of conditioning trials. **C** glu/GABA ratio did not change during presentation of the visual cue or outcome, relative to the ITI period ('Visual cue'-'ITI': t_{18}=-0.18, p=0.862; 'Outcome' - 'ITI': t_{18}=-0.32, p=0.754).



Supplementary Figure 6.11 Before and after memory recall, the hippocampal BOLD signal did not positively predict glu/GABA ratio in V1

A-B Control analyses for Figure 6.5D; MNI coordinates. Upper left: ROIs in parietal-occipital cortex (*A*) and brain stem (*B*), defined using a 12 mm sphere. Note: these control regions were restricted to the partial epi volume shown in Figure 6.3A. Upper right: In these two brain regions, the difference in BOLD signal for 'remembered' versus 'forgotten' trials did not significantly predict the increase in glu/GABA ratio observed in V1 (Figure 6.4B,F) (*A*, parietal-occipital cortex: r_{15}=0.14, p=0.592; *B*, brainstem: r_{15}=0.28, p=0.281). Lower: Moving average showing the ratio of 'remembered' to 'forgotten' trials during the inference test: BOLD signal from parietal-occipital cortex (*A*, range [-10:10]) and brain stem (*B*, range [-4:4]) shown in grey, glutamate:tCr (red, range [-8:8]), GABA:tCr (blue, range [-15:15]). Each point represents a 2.5s time bin (mean ± SEM). By incorporating a random jitter in the behavioural paradigm, MRS spectra across all trials and all participants were acquired in all possible 2.5 s time bins of the inference test trial (Supplementary Figure 6.3), thus achieving a higher temporal resolution than the TR of 4

s (see *Methods*). **C-D** During 'remembered' relative to 'forgotten' trials on the inference test (the question period, Figure 6.1D) we observed a significantly positive correlation between the hippocampal BOLD signal and glu/GABA ratio in visual cortex (Figure 6.5B-C). To assess how transient this relationship was, we assessed the relationship between hippocampal BOLD and neocortical glu/GABA ratio during the period immediately before (*C*) and after the question period (*D*). **C** Immediately prior to the question period, during presentation of the auditory cue ('Tone'), the change in hippocampal BOLD signal between subsequently 'remembered' and 'forgotten' trials did not predict the equivalent change in glu/GABA ratio in V1. However, during the auditory cue a trend towards a positive relationship was observed between the hippocampal BOLD signal and glutamate ratio in V1 (glutamate:tCr: r_{15}=0.427, p=0.089; GABA:tCr: r_{15}=0.20, p=0.438; glu/GABA ratio: r_{15}=-0.05, p=0.846; hippocampal ROI as shown in Figure 6.5A). **D** Immediately after the question period, during the inter-trial interval ('ITI'), the change in hippocampal BOLD signal between subsequently remembered and forgotten trials did not predict the equivalent change in glu/GABA ratio in V1 (glutamate:tCr: r_{15}=-0.21, p=0.421; GABA:tCr: r_{15}=-0.02, p=0.951; glu/GABA ratio: r_{15}=-0.05, p=0.854; hippocampal ROI as shown in Figure 6.5A).

6.8 Supplementary Tables

Supplementary Table 6.1 The effect of sex on behaviour and on glu/GABA ratio in V1
Using a GLM, differences in sex (male or female) were regressed onto behavioural performance during both the inference test and associative test, and onto glu/GABA ratio during the question period of the inference test. No significant effect of sex was observed.

Test	T-statistic	p-value
Behavioural performance during the *inference test* (performed inside MRI scanner)	t_{17}=0.50	p=0.622
Behavioural performance during the post-scan *associative test*	t_{17}=0.42	p=0.674
glu/GABA ratio in V1 during the question period in the *inference test*	t_{16}=1.67	p=0.215

Supplementary Table 6.2 Number of trials per condition
The number of trials per condition, reported as mean ± SEM.

Categorisation criteria	**No. of trials per condition**	
	'Remembered'	**'Forgotten'**
Using performance on both the inference test and post-scan associative test: 'correctly inferred & recalled' vs 'incorrectly inferred | not recalled' (see Figure 6.3B)	38.95 ± 1.56	39.32 ± 1.61
Using performance on inference test alone: 'correctly inferred' vs 'incorrectly inferred'	59.74 ± 1.07	18.53 ± 1.04
Using performance on post-scan associative test alone: 'recalled' vs 'not recalled'	42.84 ±1.46	25.53 ±1.51

Supplementary Table 6.3 Number of trials split according to outcome
The number of inference trials per condition ('remembered' and 'forgotten', see *Methods* for definition), split according to whether the auditory cue was indirectly associated with a rewarding or neutral outcome (set 1: rewarded; set 2: neutral), reported as mean ± SEM. There was no significant difference in the number of trials split by set (memory x set, two-way ANOVA: $F_{(1,68)}$=0.67, p=0.424). Notably, the average difference in the total number of trials in set 1 and 2 was less than 1 trial, suggesting memory recall was not confounded by reward status.

Condition	**No. of trials in inference test**	
	Set 1 (rewarded)	**Set 2 (neutral)**
'Remembered'	18.0 ± 1.15	17.33 ± 0.91
'Forgotten'	15.83 ± 0.89	16.72 ± 0.90

Supplementary Table 6.4 fMRI contrast for 'remembered' – 'forgotten'

The fMRI BOLD signal was assessed for a contrast comparing 'remembered' and 'forgotten' trials (Figure 6.3B) in the inference test. Brain regions that survived whole-volume correction for multiple comparisons are listed ($p<0.05$ with whole-brain FWE correction at the cluster-level). MNI coordinates are listed. Notably, the activation in hippocampus was bilateral but the peak on the right-hand side was not significant when applying whole-brain FWE correction ($t_{17}=3.74$, $p>0.05$, [x=22, y=-24, z=-13]).

Brain region	$P_{FWE\text{-}corr,\ cluster\ level}$	T	Peak coordinate in cluster		
			x	**y**	**z**
Left hippocampus	P=0.017	4.33	-18	-28	-8
Visual cortex	P<0.001	5.92	4	-78	0
Right auditory cortex	P<0.001	5.58	38	-26	16
Left posterior parietal cortex	P<0.001	5.00	-44	-52	22

Supplementary Table 6.5 Inter-subject covariance of glutamate and GABA

Inter-subject covariances (%) for the key metabolite measurements during the 'Question' period of inference trials (presented in Figure 6.4D-E).

	glutamate	GABA
'Remembered'	6.71	33.25
'Forgotten'	5.69	31.04

Supplementary Table 6.6 Average number of spectra (NEX)

The number of spectra contributing to metabolite estimates during the various trial periods in the inference test (mean ± SEM).

	Tone	Question	ITI
'Remembered'	60.05±2.73	16.17±1.08	35.53±1.76
'Forgotten'	47.42±2.23	14.61±1.21	28.37±1.49

Supplementary Table 6.7 Covariance between hippocampal BOLD signal and fMRS for remembered vs. forgotten

The relationship between fMRI and fMRS during 'remembered' versus 'forgotten' trials in the inference test was assessed. To this end, fMRS measures of glu/GABA ratio from V1 for 'remembered' – 'forgotten' were included as covariates in a group analysis for the equivalent fMRI contrast ($p<0.05$ with FWE correction at the cluster-level). The only brain region to survive whole-brain correction for multiple comparisons was the left hippocampus. Thus, the BOLD signal in left hippocampus significantly predicted individual differences in glu/GABA ratio measured from V1 during memory recall. MNI coordinates.

Brain region	$P_{FWE\text{-}corr,\ peak\ level}$	T	Coordinate		
			x	**y**	**z**
Left hippocampus	P=0.005	11.37	-26	-12	-16

Supplementary Table 6.8 Minimum Reporting Standards in MRS (MRSinMRS) checklist

Site (Name or Number) 1. Hardware	FMRIB, Wellcome Centre for Integrative Neuroimaging (University of Oxford)
a. Field strength [T]	7
b. Manufacturer	Siemens
c. Model (software version if available)	Magnetom 7T (VB17a)
d. RF coils: nuclei (transmit/ receive), number of channels, type, body part	1-channel transmit and a 32-channel receive phased-array head coil (Nova Medical Inc., USA)
e. Additional hardware	Two 110 × 110 × 5 mm^3 Barium Titanate dielectric pads (4:1 ratio of $BaTiO_3$:D_2O, relative permittivity around 300) over occipital lobe
2. Acquisition	
a. Pulse sequence	combined fMRI-fMRS sequence fMRI: 3D EPI fMRS: semi-LASER
b. Volume of Interest (VOI) locations	Visual cortex
c. Nominal VOI size [cm^3, mm^3]	2 × 2 × 2 cm^3
d. Repetition Time (TR), Echo Time (TE) [ms, s]	TR = 4105 ms TE = 36 ms
e. Total number of Excitations or acquisitions per spectrum In time series for kinetic studies i. Number of Averaged spectra (NA) per time-point ii. Averaging method (e.g. block-wise or moving average) iii. Total number of spectra (acquired / in time-series)	Average total number of spectra acquired per participant: N=457 (SD: 35.62) Between 8 and 81 spectra per condition (see Supplementary Table 6.6). Participants with less than 8 spectra per condition (e.g. 'remembered' or 'forgotten') were not included in the analysis. Spectra were averaged according to condition ('remembered'/'forgotten').
f. Additional sequence parameters (spectral width in Hz, number of spectral points, frequency offsets) If STEAM:, Mixing Time (TM) If MRSI: 2D or 3D, FOV in all directions, matrix size, acceleration factors, sampling method	Spectral width: 6 kHz Spectral points: 2048 frequency offset: -2.0 ppm
g. Water Suppression Method	VAPOR
h. Shimming Method, reference peak, and thresholds for "acceptance of shim" chosen	3D head shim (GRESHIM) followed by FASTESTMAP, threshold for acceptable shim was linewidth water <16 Hz
i. Triggering or motion correction method (respiratory, peripheral, cardiac triggering, incl. device used and delays)	N/A
3. Data analysis methods and outputs	
a. Analysis software	LCmodel (fitting and quantification, version 6.3-1B), MRspa (preprocessing, version 1.4)
b. Processing steps deviating from quoted reference or product	Residual water component was removed using Hankel Lanczos Singular Value Decomposition (HLSVD).
c. Output measure (e.g. absolute concentration, institutional units, ratio)Processing steps deviating from quoted reference or product	Concentrations relative to total Creatine.
d. Quantification references and assumptions, fitting model assumptions	The basis set included simulated model spectra of alanine, aspartate, ascorbate/vitamin C, glycerophosphocholine, phosphocholine, creatine, phosphocreatine, GABA, glucose, glutamine,

	glutamate, glutathione, myo-inositol, Lactate, N-acetylaspartate, N-acetylaspartylglutamate, phosphoethanolamine, scyllo-inositol, taurine and experimentally measured macromolecules. Basis set is made available on GitHub. Prior constraints that fix GABA estimates relative to more abundant neurochemicals were turned off (i.e. NRATIO set to 0).
4. Data Quality	
a. Reported variables (SNR, Linewidth (with reference peaks))	Across all spectra (Supplementary Figure 6.5): SNR: 51.1± 2.37% (mean ± SEM) FWHM from LCModel: 9.86±0.30 Hz (mean ± SEM) Total Creatine Linewidth: 11.01±0.32Hz (mean ± SEM)
b. Data exclusion criteria	Subjects were excluded from an analysis if there were less than 8 spectra acquired for one or more of the conditions of interest (n=1 subject excluded for analyses concerning the question period due to less than 8 spectra in the 'forgotten' condition).
c. Quality measures of postprocessing Model fitting (e.g. CRLB, goodness of fit, SD of residual)	Across all spectra (Supplementary Figure 6.5, mean ± SEM): CRLB glutamate: 4.42±0.43% CRLB GABA: 11.37±0.94% Intra-subject covariance glutamate: 2.68±0.62% Intra-subject covariance GABA was 8.74±1.95% Inter-subject covariance (Supplementary Table 6.4): glutamate: 5.69-6.71% GABA: 31.04-33.25%
d. Sample Spectrum	Figure 6.3E

CHAPTER 7

GENERAL DISCUSSION AND FUTURE DIRECTIONS

The brain has the remarkable capacity to acquire, store, and reinstate memories. Central to understanding these memory-related processes is the trade-off between flexibility and stability. Memories need to be stored in a stable manner, safe from interference, to allow accurate recall. On the other hand, to form new memories and to edit memories in accordance with new experiences, plasticity is needed.

In this thesis, I have investigated the neural mechanisms that support stable and flexible memory, probing neocortical, hippocampal and neuromodulatory contributions. To disentangle these different contributions, I have used techniques that measure and modulate neural excitation and inhibition in combination with behavioural and neural measures of memory representations. Through these approaches my aim was to understand how local neurochemistry may influence memory related processes. In Chapter 4, I investigated how during memory formation, E/I balance may be regulated by neuromodulators to set the trade-off between stable and flexible memory. In Chapter 5, I asked what mechanisms protect overlapping memories from interfering with each other. In Chapter 6, I investigated how memory-supporting mechanisms in hippocampus and neocortex may interact to facilitate memory recall. Overall, I found that E/I balance plays a governing role in all memory processes investigated in this thesis. This finding illustrates the relevance and importance of considering neurochemistry when trying to understand cognitive processing. In this final chapter, I will briefly discuss some of the general implications of these findings on our understanding of memory and learning, and discuss possible future directions that may be of interest.

7.1 Revealing the role of E/I balance in memory

A focus of the work presented in this thesis has been to reveal how E/I balance, and specifically neural inhibition, supports memory-related processes in the brain. Neural inhibition appears to be critical in gating the re-expression of stored memories (Chapter 5 and 6). The results from these chapters support the hypothesis that memories are reactivated through disinhibition, releasing them from a blanket of inhibition (Karnani et al., 2016). This hypothesis sheds an interesting light

on the concept of 'forgetting', which we typically conceptualise as information being lost, i.e. no longer stored. However, it may simply be that this information is no longer accessible (Frankland et al., 2019; Tulving, 1974), due to the memory being masked by inhibition. An interesting question to address in humans would therefore be whether it is possible to reactivate forgotten memories by modulating E/I balance during recall. Here, lowering inhibition could facilitate the reactivation of memory traces that are otherwise no longer accessible.

Throughout the work presented in this thesis, I have used MRS-derived concentrations of glutamate and GABA, where glutamate has been a proxy for excitatory tone and GABA a proxy of inhibitory tone. In Chapter 4 I modulated GABA using a pharmacological intervention and in Chapter 5 I lowered cortical GABA using anodal tDCS to reveal the role of E/I balance in learning and memory storage, respectively. In Chapter 6 I evaluated GABA and glutamate in a time-resolved manner and found a reduction in GABA during successful retrieval of a memory. The benefit of approximating excitation and inhibition using MRS is that this approach can be applied non-invasively in humans. However, the spatiotemporal resolution of MRS-derived glutamate and GABA is limited, and it does not allow for dissociating the contribution of different inhibitory interneuron subtypes. Furthermore, as discussed in Chapters 3 and 4, interpretation of MRS-derived glutamate and GABA concentrations is not straightforward, as MRS-derived measurements of neural metabolites do not directly equate to excitatory and inhibitory potentials, but instead may reflect shifts in metabolism. Thus, the research in humans as I have presented here is limited in revealing the underlying circuit-level mechanisms of memory. Other than intracranial recordings in patients, which remain rare, invasive research in animal models is needed to provide insight into microcellular neural processes and connectivity. Here, cross-species research provides the opportunity to link research in animal models and humans, with the potential to translate microcircuit findings to higher-level cognition (Barron et al., 2021).

7.2 Interactions between brain regions that support memory

One of the questions that I aimed to address is how hippocampus and neocortex together support memory functions like learning, storage and recall. Overall, I have taken a correlational approach and observed that activity in hippocampus is related to cortical E/I balance in the visual cortex immediately following learning (Chapter 4), when neural inhibition is not perturbed using tDCS (Chapter 5), and during successful recall (Chapter 6). However, these findings leave open the nature of the information shared between these brain regions. An interesting possibility is that (aspects of) cognitive maps may be shared between brain regions (Morton et al., 2017; Wikenheiser

and Schoenbaum, 2016), and that this process could occur over time in accordance with complementary learning systems theory (McClelland et al., 1995).

In relation to this, in Chapter 4 I found that neural excitability during learning affects behavioural read-outs of the cognitive map formed. These behavioural findings therefore suggest that there was an effect of inhibition on how memories were consolidated. Since memory replay is thought to underlie memory consolidation (Louie and Wilson, 2001; Wilson and McNaughton, 1994), it would be interesting to investigate whether E/I balance during learning modulates replay events following learning, and whether replay in turn could predict the behavioural distortions as observed in Chapter 4.

The work presented in this thesis leaves open the role of other cortical areas and brain regions in general in memory-related processes. It would be particularly interesting to address the role of the prefrontal cortex (PFC) during memory recall, given its suggested importance in longer-term memory consolidation (Preston and Eichenbaum, 2013), and given the recent finding that PFC may assert control over hippocampus through long-range inhibitory projections (Malik et al., 2022). Could PFC activity be critical in reactivating the context-relevant neural representations in hippocampus and other cortical areas? In humans, this question could potentially be addressed using neuromodulatory techniques that appear to enhance or disrupt information transfer or connectivity between regions, such as transcranial Alternating Current Stimulation (tACS), or Transcranial Ultrasonic Stimulation (TUS) (Bächinger et al., 2017; Folloni et al., 2019; Verhagen et al., 2019; Weinrich et al., 2017).

7.3 Concluding remarks

In this thesis, I have presented three studies that address the neural mechanisms underlying the three processes by which we typically define memory, namely memory formation, storage and recall. Using a combination of neuromodulatory approaches, measures of functional brain activity and neurochemistry, I was able to reveal the roles E/I balance and hippocampus play in the lifetime of a memory. Taken together, these studies highlight that E/I balance is involved in memory formation, storage and recall, where neural inhibition appears to protect memories from overgeneralisation and interference, and where disinhibition releases memories to be reactivated. By setting the balance between stability and flexibility, E/I balance may play a key role in neural computations that underlie flexible behaviour.

BIBLIOGRPAHY

Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B., 1997. Synaptic depression and cortical gain control. Science 275, 220–224. https://doi.org/10.1126/science.275.5297.221

Albrecht, D.G., Farrar, S.B., Hamilton, D.B., 1984. Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. The Journal of Physiology 347, 713–739. https://doi.org/10.1113/jphysiol.1984.sp015092

Amir, Y., Harel, M., Malach, R., 1993. Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. J Comp Neurol 334, 19–46. https://doi.org/10.1002/cne.903340103

Andersen, P., Bliss, T.V.P., Skrede, K.K., 1971. Lamellar organization of hippocampal excitatory pathways. Exp Brain Res 13, 222–238. https://doi.org/10.1007/BF00234087

Anderson, J.S., Carandini, M., Ferster, D., 2000. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. J Neurophysiol 84, 909–926. https://doi.org/10.1152/jn.2000.84.2.909

Andreychenko, A., Boer, V.O., Castro, C.S.A. de, Luijten, P.R., Klomp, D.W.J., 2012. Efficient spectral editing at 7 T: GABA detection with MEGA-sLASER. Magnetic Resonance in Medicine 68, 1018–1025. https://doi.org/10.1002/mrm.24131

Apšvalka, D., Gadie, A., Clemence, M., Mullins, P.G., 2015. Event-related dynamics of glutamate and BOLD effects measured using functional magnetic resonance spectroscopy (fMRS) at 3 T in a repetition suppression paradigm. NeuroImage 118, 292–300. https://doi.org/10.1016/j.neuroimage.2015.06.015

Arolfo, M.P., Brioni, J.D., 1991. Diazepam impairs place learning in the Morris water maze. Behav Neural Biol 55, 131–136. https://doi.org/10.1016/0163-1047(91)80133-y

Aronov, D., Nevers, R., Tank, D.W., 2017. Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. Nature 543, 719–722. https://doi.org/10.1038/nature21692

Aston-Jones, G., Bloom, F.E., 1981. Norepinephrine-containing locus coeruleus neurons in behaving rats exhibit pronounced responses to non-noxious environmental stimuli. J Neurosci 1, 887–900.

Aston-Jones, G., Cohen, J.D., 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu Rev Neurosci 28, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bächinger, M., Zerbi, V., Moisa, M., Polania, R., Liu, Q., Mantini, D., Ruff, C., Wenderoth, N., 2017. Concurrent tACS-fMRI Reveals Causal Influence of Power Synchronized Neural Activity on Resting State fMRI Connectivity. J. Neurosci. 37, 4766–4777. https://doi.org/10.1523/jneurosci.1756-16.2017

Bak, L.K., Schousboe, A., Waagepetersen, H.S., 2006. The glutamate/GABA-glutamine cycle: aspects of transport, neurotransmitter homeostasis and ammonia transfer. Journal of Neurochemistry 98, 641–653. https://doi.org/10.1111/j.1471-4159.2006.03913.x

Baker, C.I., Behrmann, M., Olson, C.R., 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nature Neuroscience 5, 1210–1216. https://doi.org/10.1038/nn960

Bakker, A., Kirwan, C.B., Miller, M., Stark, C.E.L., 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. Science 319, 1640–1642. https://doi.org/10.1126/science.1152882

Barron, H.C.*, Garvert, M.M.*, Behrens, T.E.J., 2016b. Repetition suppression: a means to index neural representations using BOLD? Philosophical Transactions of the Royal Society B: Biological Sciences 371, 20150355. https://doi.org/10.1098/rstb.2015.0355

Barron, H.C., Mars, R.B., Dupret, D., Lerch, J.P., Sampaio-Baptista, C., 2021. Cross-species neuroscience: closing the explanatory gap. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20190633. https://doi.org/10.1098/rstb.2019.0633

Barron, H.C., Reeve, H.M., Koolschijn, R.S., Perestenko, P.V., Shpektor, A., Nili, H., Rothaermel, R., Campo-Urriza, N., O'Reilly, J.X., Bannerman, D.M., Behrens, T.E.J., Dupret, D., 2020. Neuronal Computation Underlying Inferential Reasoning in Humans and Mice. Cell. https://doi.org/10.1016/j.cell.2020.08.035

Barron, H.C., Vogels, T.P., Behrens, T.E., Ramaswami, M., 2017. Inhibitory engrams in perception and memory. Proc Natl Acad Sci U S A 114, 6666–6674. https://doi.org/10.1073/pnas.1701812114

Barron, H.C., Vogels, T.P., Emir, U.E., Makin, T.R., O'Shea, J., Clare, S., Jbabdi, S., Dolan, R.J., Behrens, T.E.J., 2016a. Unmasking Latent Inhibitory Connections in Human Cortex to Reveal Dormant Cortical Memories. Neuron 90, 191–203. https://doi.org/10.1016/j.neuron.2016.02.031

Bartfeld, E., Grinvald, A., 1992. Relationships between orientation-preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex. Proceedings of the National Academy of Sciences 89, 11905–11909. https://doi.org/10.1073/pnas.89.24.11905

Baylis, G.C., Rolls, E.T., 1987. Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp Brain Res 65, 614–622. https://doi.org/10.1007/BF00235984

Bednařík, P., Tkáč, I., Giove, F., DiNuzzo, M., Deelchand, D.K., Emir, U.E., Eberly, L.E., Mangia, S., 2015. Neurochemical and BOLD responses during neuronal activation measured in the human visual cortex at 7 Tesla. Journal of Cerebral Blood Flow and Metabolism 35, 601–610.

Bednařík, P., Tkáč, I., Giove, F., Eberly, L.E., Deelchand, D.K., Barreto, F.R., Mangia, S., 2018. Neurochemical responses to chromatic and achromatic stimuli in the human visual cortex. J Cereb Blood Flow Metab 38, 347–359. https://doi.org/10.1177/0271678X17695291

Befroy, D.E., Shulman, G.I., 2011. Magnetic Resonance Spectroscopy Studies of Human Metabolism. Diabetes 60, 1361–1369. https://doi.org/10.2337/db09-0916

Bellmund, J.L.S., de Cothi, W., Ruiter, T.A., Nau, M., Barry, C., Doeller, C.F., 2020. Deforming the metric of cognitive maps distorts memory. Nat Hum Behav 4, 177–188. https://doi.org/10.1038/s41562-019-0767-3

Benchenane, K., Peyrache, A., Khamassi, M., Tierney, P.L., Gioanni, Y., Battaglia, F.P., Wiener, S.I., 2010. Coherent theta oscillations and reorganization of spike timing in the hippocampal- prefrontal network upon learning. Neuron 66, 921–936. https://doi.org/10.1016/j.neuron.2010.05.013

Bergado-Acosta, J.R., Sangha, S., Narayanan, R.T., Obata, K., Pape, H.-C., Stork, O., 2008. Critical role of the 65-kDa isoform of glutamic acid decarboxylase in consolidation and generalization of Pavlovian fear memory. Learn Mem 15, 163–171. https://doi.org/10.1101/lm.705408

Berger, T.K., Perin, R., Silberberg, G., Markram, H., 2009. Frequency-dependent disynaptic inhibition in the pyramidal network: a ubiquitous pathway in the developing rat neocortex. The Journal of Physiology 587, 5411–5425. https://doi.org/10.1113/jphysiol.2009.176552

Berridge, C.W., 2008. Noradrenergic modulation of arousal. Brain Res Rev 58, 1–17. https://doi.org/10.1016/j.brainresrev.2007.10.013

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H.J., Kumaran, D., Düzel, E., 2016. Strong Evidence for Pattern Separation in Human Dentate Gyrus. J. Neurosci. 36, 7569–7579. https://doi.org/10.1523/JNEUROSCI.0518-16.2016

Bilodeau, I.M., Schlosberg, H., 1951. Similarity in stimulating conditions as a variable in retroactive inhibition. Journal of Experimental Psychology 41, 199–204. https://doi.org/10.1037/h0056809
Bindman, L.J., Lippold, O.C.J., Redfearn, J.W.T., 1962. Long-lasting Changes in the Level of the Electrical Activity of the Cerebral Cortex produced by Polarizing Currents. Nature 196, 584–585. https://doi.org/10.1038/196584a0
Blakemore, C., Campbell, F.W., 1969. Adaptation to spatial stimuli. The Journal of physiology 200, 11P-13P.
Bliss, T.V., Lomo, T., 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 232, 331–356. https://doi.org/10.1113/jphysiol.1973.sp010273
Bliss, T.V.P., Collingridge, G.L., 1993. A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361, 31–39. https://doi.org/10.1038/361031a0
Bonnici, H.M., Kumaran, D., Chadwick, M.J., Weiskopf, N., Hassabis, D., Maguire, E.A., 2012. Decoding representations of scenes in the medial temporal lobes. Hippocampus 22, 1143–1153. https://doi.org/10.1002/hipo.20960
Boorman, E.D., Rajendran, V.G., O'Reilly, J.X., Behrens, T.E., 2016. Two Anatomically and Computationally Distinct Learning Signals Predict Changes to Stimulus-Outcome Associations in Hippocampus. Neuron 89, 1343–1354. https://doi.org/10.1016/j.neuron.2016.02.014
Bosch, S.E., Jehee, J.F.M., Fernández, G., Doeller, C.F., 2014. Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. J Neurosci 34, 7493–7500. https://doi.org/10.1523/JNEUROSCI.0805-14.2014
Bottomley, P.A., 1987. Spatial Localization in NMR Spectroscopy in Vivo. Ann NY Acad Sci 508, 333–348. https://doi.org/10.1111/j.1749-6632.1987.tb32915.x
Bouret, S., Sara, S.J., 2005. Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. Trends in Neurosciences 28, 574–582. https://doi.org/10.1016/j.tins.2005.09.002
Brink, W.M., Webb, A.G., 2014. High Permittivity Pads Reduce Specific Absorption Rate , Improve B1 Homogeneity, and Increase Contrast-to-Noise Ratio for Functional Cardiac MRI at 3 T. Magnetic Resonance in Medicine 71, 1632–1640. https://doi.org/10.1002/mrm.24778
Brioni, J.D., Nagahara, A.H., McGaugh, J.L., 1989. Involvement of the amygdala GABAergic system in the modulation of memory storage. Brain Research 487, 105–112. https://doi.org/10.1016/0006-8993(89)90945-1
Brogden, W.J., 1939. Sensory pre-conditioning. Journal of Experimental Psychology 25, 323–332. https://doi.org/10.1037/h0058944
Brunelin, J., Mondino, M., Gassab, L., Haesebaert, F., Gaha, L., Suaud-Chagny, M.-F., Saoud, M., Mechri, A., Poulet, E., 2012. Examining Transcranial Direct-Current Stimulation (tDCS) as a Treatment for Hallucinations in Schizophrenia. AJP 169, 719–724. https://doi.org/10.1176/appi.ajp.2012.11071091
Bulkin, D.A., Law, L.M., Smith, D.M., 2016. Placing memories in context: Hippocampal representations promote retrieval of appropriate memories. Hippocampus 26, 958–971. https://doi.org/10.1002/hipo.22579
Bunsey, M., Eichenbaum, H., 1996. Conservation of hippocampal memory function in rats and humans. Nature 379, 255–257.
Busche, M.A., Konnerth, A., 2016. Impairments of neural circuit function in Alzheimer's disease. Philos Trans R Soc Lond B Biol Sci 371, 20150429. https://doi.org/10.1098/rstb.2015.0429
Bush, D., Barry, C., Manson, D., Burgess, N., 2015. Using Grid Cells for Navigation. Neuron 87, 507–520. https://doi.org/10.1016/j.neuron.2015.07.006

Bütefisch, C.M., Davis, B.C., Wise, S.P., Sawaki, L., Kopylev, L., Classen, J., Cohen, L.G., 2000. Mechanisms of use-dependent plasticity in the human motor cortex. Proc Natl Acad Sci U S A 97, 3661–3665. https://doi.org/10.1073/pnas.97.7.3661

Butler, W.N., Hardcastle, K., Giocomo, L.M., 2019. Remembered reward locations restructure entorhinal spatial maps. Science 363, 1447–1452. https://doi.org/10.1126/science.aav5297

Butterly, D.A., Petroccione, M.A., Smith, D.M., 2012. Hippocampal context processing is critical for interference free recall of odor memories in rats. Hippocampus 22, 906–913. https://doi.org/10.1002/hipo.20953

Buzsáki, G., Moser, E.I., 2013. Memory, navigation and theta rhythm in the hippocampal-entorhinal system. Nat Neurosci 16, 130–138. https://doi.org/10.1038/nn.3304

Bymaster, F.P., Katner, J.S., Nelson, D.L., Hemrick-Luecke, S.K., Threlkeld, P.G., Heiligenstein, J.H., Morin, S.M., Gehlert, D.R., Perry, K.W., 2002. Atomoxetine increases extracellular levels of norepinephrine and dopamine in prefrontal cortex of rat: a potential mechanism for efficacy in attention deficit/hyperactivity disorder. Neuropsychopharmacology 27, 699–711. https://doi.org/10.1016/S0893-133X(02)00346-9

Cabanes, E., Confort-Gouny, S., Le Fur, Y., Simond, G., Cozzone, P.J., 2001. Optimization of Residual Water Signal Removal by HLSVD on Simulated Short Echo Time Proton MR Spectra of the Human Brain. Journal of Magnetic Resonance 150, 116–125. https://doi.org/10.1006/jmre.2001.2318

Cahill, L., McGaugh, J.L., Weinberger, N.M., 2001. The neurobiology of learning and memory: some reminders to remember. Trends in Neurosciences 24, 578–581. https://doi.org/10.1016/S0166-2236(00)01885-3

Cai, D.J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S.E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., Golshani, P., Silva, A.J., 2016. A shared neural ensemble links distinct contextual memories encoded close in time. Nature 534, 115–118. https://doi.org/10.1038/nature17955

Canto, C.B., Wouterlood, F.G., Witter, M.P., 2008. What does anatomical organization of entorhinal cortex tell us? Neural Plasticity 2008. https://doi.org/10.1155/2008/381243

Cao, P., Wu, E.X., 2017. In vivo diffusion MRS investigation of non-water molecules in biological tissues. NMR in Biomedicine 30, e3481. https://doi.org/10.1002/nbm.3481

Carandini, M., Ferster, D., 1997. A Tonic Hyperpolarization Underlying Contrast Adaptation in Cat Visual Cortex. Science 276, 949–952. https://doi.org/10.1126/science.276.5314.949

Carpenter, A.C., Schacter, D.L., 2017. Flexible retrieval: When true inferences produce false memories. Journal of Experimental Psychology: Learning, Memory, and Cognition 43, 335–349. https://doi.org/10.1037/xlm0000340

Carvalho, T.P., Buonomano, D.V., 2009. Differential Effects of Excitatory and Inhibitory Plasticity on Synaptically Driven Neuronal Input-Output Functions. Neuron 61, 774–785. https://doi.org/10.1016/j.neuron.2009.01.013

Castro-Alamancos, M.A., Donoghue, J.P., Connors, B.W., 1995. Different forms of synaptic plasticity in somatosensory and motor areas of the neocortex. J. Neurosci. 15, 5324–5333. https://doi.org/10.1523/jneurosci.15-07-05324.1995

Chamberland, S., Topolnik, L., 2012. Inhibitory control of hippocampal inhibitory neurons. Frontiers in Neuroscience 6.

Chandler, D., Waterhouse, B., 2012. Evidence for Broad Versus Segregated Projections from Cholinergic and Noradrenergic Nuclei to Functionally and Anatomically Discrete Subregions of Prefrontal Cortex. Frontiers in Behavioral Neuroscience 6.

Chandler, D.J., Gao, W.-J., Waterhouse, B.D., 2014. Heterogeneous organization of the locus coeruleus projections to prefrontal and motor cortices. Proc Natl Acad Sci U S A 111, 6816–6821. https://doi.org/10.1073/pnas.1320827111

Chen, C., Morris, P., Francis, S., Gowland, P., 2015. A comparison of MEGA-sLASER and STEAM for in vivo quantification of GABA at 7T, in: Proc Int Soc Magn Reson Med.
Chen, J., Olsen, R.K., Preston, A.R., Glover, G.H., Wagner, A.D., 2011. Associative retrieval processes in the human medial temporal lobe: Hippocampal retrieval success and CA1 mismatch detection. Learn. Mem. 18, 523–528. https://doi.org/10.1101/lm.2135211
Chen, N., Sugihara, H., Sur, M., 2015. An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. Nat Neurosci 18, 892–902. https://doi.org/10.1038/nn.4002
Chhatwal, J.P., Myers, K.M., Ressler, K.J., Davis, M., 2005. Regulation of Gephyrin and GABAA Receptor Binding within the Amygdala after Fear Acquisition and Extinction. J. Neurosci. 25, 502–506. https://doi.org/10.1523/jneurosci.3301-04.2005
Clarke, W.T., Stagg, C.J., Jbabdi, S., 2021. FSL-MRS: An end-to-end spectroscopy analysis package. Magnetic Resonance in Medicine 85, 2950–2964. https://doi.org/10.1002/mrm.28630
Cleve, M., Gussew, A., Reichenbach, J.R., 2015. In vivo detection of acute pain-induced changes of GABA+ and Glx in the human brain by using functional 1H MEGA-PRESS MR spectroscopy. NeuroImage 105, 67–75. https://doi.org/10.1016/j.neuroimage.2014.10.042
Clopath, C., Vogels, T.P., Froemke, R.C., Sprekeler, H., 2016. Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. https://doi.org/10.1101/066589
Cohen, N.J., Eichenbaum, H., 1993. Memory, Amnesia, and the Hippocampal System. A Bradford Book, Cambridge, MA, USA.
Collingridge, G.L., Kehl, S.J., McLennan, H., 1983. Excitatory amino acids in synaptic transmission in the Schaffer collateral-commissural pathway of the rat hippocampus. The Journal of Physiology 334, 33–46. https://doi.org/10.1113/jphysiol.1983.sp014478
Constantinescu, A.O.*, O'Reilly, J.X.*, Behrens, T.E.J., 2016. Organizing conceptual knowledge in humans with a gridlike code. Science 352, 1464–1468. https://doi.org/10.1126/science.aaf0941
Couey, J.J., Witoelar, A., Zhang, S.-J., Zheng, K., Ye, J., Dunn, B., Czajkowski, R., Moser, M.-B., Moser, E.I., Roudi, Y., Witter, M.P., 2013. Recurrent inhibitory circuitry as a mechanism for grid formation. Nat Neurosci 16, 318–324. https://doi.org/10.1038/nn.3310
Courtin, J., Chaudun, F., Rozeske, R.R., Karalis, N., Gonzalez-Campo, C., Wurtz, H., Abdi, A., Baufreton, J., Bienvenu, T.C.M., Herry, C., 2014. Prefrontal parvalbumin interneurons shape neuronal activity to drive fear expression. Nature 505, 92–96. https://doi.org/10.1038/nature12755
Cox, D.D., Savoy, R.L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19, 261–270. https://doi.org/10.1016/S1053-8119(03)00049-1
D'amour, J.A., Froemke, R.C., 2015. Inhibitory and Excitatory Spike-Timing-Dependent Plasticity in the Auditory Cortex. Neuron 86, 514–528. https://doi.org/10.1016/j.neuron.2015.03.014
Dayan, P., 1993. Improving generalization for temporal difference learning: The successor representation. Neural Computation 5, 613–624. https://doi.org/10.1162/neco.1993.5.4.613
de Almeida, L., Idiart, M., Lisman, J.E., 2009. A Second Function of Gamma Frequency Oscillations: An E%-Max Winner-Take-All Mechanism Selects Which Cells Fire. J. Neurosci. 29, 7497–7503. https://doi.org/10.1523/jneurosci.6044-08.2009
De Graaf, A.A., Bovée, W.M.M.J., 1990. Improved quantification of in vivo1H NMR spectra by optimization of signal acquisition and processing and by incorporation of prior knowledge into the spectral fitting. Magnetic Resonance in Medicine 15, 305–319. https://doi.org/10.1002/mrm.1910150212

Depue, B.E., Curran, T., Banich, M.T., 2007. Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. Science 317, 215–219. https://doi.org/10.1126/science.1139560
DeVito, L.M., Kanter, B.R., Eichenbaum, H., 2010. The hippocampus contributes to memory expression during transitive inference in mice. Hippocampus 20, 208–217.
Diedrichsen, J., Kriegeskorte, N., 2017. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLOS Computational Biology 13, e1005508. https://doi.org/10.1371/journal.pcbi.1005508
Ding, Y.-S., Naganawa, M., Gallezot, J.-D., Nabulsi, N., Lin, S.-F., Ropchan, J., Weinzimmer, D., McCarthy, T.J., Carson, R.E., Huang, Y., Laruelle, M., 2014. Clinical doses of atomoxetine significantly occupy both norepinephrine and serotonin transports: Implications on treatment of depression and ADHD. Neuroimage 86, 164–171. https://doi.org/10.1016/j.neuroimage.2013.08.001
Doeller, C.F., Barry, C., Burgess, N., 2010. Evidence for grid cells in a human memory network. Nature 463, 657–661. https://doi.org/10.1038/nature08704
Donato, F., Rompani, S.B., Caroni, P., 2013. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. Nature 504, 272–276. https://doi.org/10.1038/nature12866
Douglas, R.J., Martin, K.A.C., 2004. Neuronal circuits of the neocortex. Annu Rev Neurosci 27, 419–451. https://doi.org/10.1146/annurev.neuro.27.070203.144152
Dudai, A., Yayon, N., Lerner, V., Tasaka, G.-I., Deitcher, Y., Gorfine, K., Niederhoffer, N., Mizrahi, A., Soreq, H., London, M., 2020. Barrel cortex VIP/ChAT interneurons suppress sensory responses in vivo. PLoS Biol 18, e3000613. https://doi.org/10.1371/journal.pbio.3000613
Dupret, D., O'Neill, J., Pleydell-Bouverie, B., Csicsvari, J., 2010. The reorganization and reactivation of hippocampal maps predict spatial memory performance. Nat Neurosci 13, 995–1002. https://doi.org/10.1038/nn.2599
Edden, R.A.E., Puts, N.A.J., Barker, P.B., 2012. Macromolecule-suppressed GABA-edited magnetic resonance spectroscopy at 3T. Magnetic Resonance in Medicine 68, 657–661. https://doi.org/10.1002/mrm.24391
Edden, R.A.E., Puts, N.A.J., Harris, A.D., Barker, P.B., Evans, C.J., 2014. Gannet: A batch-processing tool for the quantitative analysis of gamma-aminobutyric acid–edited MR spectroscopy spectra. Journal of Magnetic Resonance Imaging 40, 1445–1452. https://doi.org/10.1002/jmri.24478
Edeline, R., J.M., Manunta, Y., Hennevin, E., 2011. Induction of selective plasticity in the frequency tuning of auditory cortex and auditory thalamus neurons by locus coeruleus stimulation. Hearing Research 274, 75–84. https://doi.org/10.1016/j.heares.2010.08.005
Efron, B., 2000. The Bootstrap and Modern Statistics. Journal of the American Statistical Association 95, 1293–1296. https://doi.org/10.1080/01621459.2000.10474333
Eggermann, E., Serafin, M., Bayer, L., Machard, D., Saint-Mleux, B., Jones, B.E., Mühlethaler, M., 2001. Orexins/hypocretins excite basal forebrain cholinergic neurones. Neuroscience 108, 177–181. https://doi.org/10.1016/s0306-4522(01)00512-7
Eichenbaum, H., 2004. Hippocampus: cognitive processes and neural representations that underlie declarative memory. Neuron 44, 109–120. https://doi.org/10.1016/j.neuron.2004.08.028
Eichenbaum, H., Cohen, N.J., 2014. Can We Reconcile the Declarative Memory and Spatial Navigation Views on Hippocampal Function? Neuron 83, 764–770. https://doi.org/10.1016/j.neuron.2014.07.032

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., Fried, I., 2003. Cellular networks underlying human spatial navigation. Nature 425, 184–188. https://doi.org/10.1038/nature01964

Eldar, E., Cohen, J.D., Niv, Y., 2013. The effects of neural gain on attention and learning. Nat Neurosci 16, 1146–1153. https://doi.org/10.1038/nn.3428

Eli Lilly and Co., 2003. Strattera™ (atomoxetine) package insert (NDA21–411). Indianapolis (IN).

Favila, S.E., Chanales, A.J.H., Kuhl, B.A., 2016. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. Nat Commun 7, 11066. https://doi.org/10.1038/ncomms11066

Fecteau, S., Knoch, D., Fregni, F., Sultani, N., Boggio, P., Pascual-Leone, A., 2007. Diminishing Risk-Taking Behavior by Modulating Activity in the Prefrontal Cortex: A Direct Current Stimulation Study. J. Neurosci. 27, 12500–12505. https://doi.org/10.1523/JNEUROSCI.3283-07.2007

Finkelman, T., Furman-Haran, E., Paz, R., Tal, A., 2022. Quantifying the excitatory-inhibitory balance: A comparison of SemiLASER and MEGA-SemiLASER for simultaneously measuring GABA and glutamate at 7T. NeuroImage 247, 118810. https://doi.org/10.1016/j.neuroimage.2021.118810

Fino, E., Yuste, R., 2011. Dense inhibitory connectivity in neocortex. Neuron 69, 1188–1203. https://doi.org/10.1016/j.neuron.2011.02.025

Fishell, G., Rudy, B., 2011. Mechanisms of Inhibition within the Telencephalon: "Where the Wild Things Are." Annual Review of Neuroscience 34, 535–567. https://doi.org/10.1146/annurev-neuro-061010-113717

Fletcher, M.L., Wilson, D.A., 2002. Experience modifies olfactory acuity: acetylcholine-dependent learning decreases behavioral generalization between similar odorants. J Neurosci 22, RC201.

Floyer-Lea, A., Wylezinska, M., Kincses, T., Matthews, P.M., 2006. Rapid modulation of GABA concentration in human sensorimotor cortex during motor learning. Journal of Neurophysiology 95, 1639–1644. https://doi.org/10.1152/jn.00346.2005

Folloni, D., Verhagen, L., Mars, R.B., Fouragnan, E., Constans, C., Aubry, J.-F., Rushworth, M.F.S., Sallet, J., 2019. Manipulation of Subcortical and Deep Cortical Activity in the Primate Brain Using Transcranial Focused Ultrasound Stimulation. Neuron 101, 1109-1116.e5. https://doi.org/10.1016/j.neuron.2019.01.019

Foote, S.L., Aston-Jones, G., Bloom, F.E., 1980. Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal. Proc Natl Acad Sci U S A 77, 3033–3037.

Fox, P.T., Raichle, M.E., 1986. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. PNAS 83, 1140–1144. https://doi.org/10.1073/pnas.83.4.1140

Fox, P.T., Raichle, M.E., Mintun, M.A., Dence, C., 1988. Nonoxidative glucose consumption during focal physiologic neural activity. Science 241, 462–464. https://doi.org/10.1126/science.3260686

Frangou, P., Emir, U.E., Karlaftis, V.M., Nettekoven, C., Hinson, E.L., Larcombe, S., Bridge, H., Stagg, C.J., Kourtzi, Z., 2019. Learning to optimize perceptual decisions through suppressive interactions in the human brain. Nat Commun 10, 474. https://doi.org/10.1038/s41467-019-08313-y

Frankland, P.W., Josselyn, S.A., Köhler, S., 2019. The neurobiological foundation of memory retrieval. Nat Neurosci 22, 1576–1585. https://doi.org/10.1038/s41593-019-0493-1

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2006. Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. Cerebral Cortex 16, 1631–1644. https://doi.org/10.1093/cercor/bhj100

Friston, K., 2005. A theory of cortical responses. Philosophical Transactions of the Royal Society B: Biological Sciences 360, 815–836. https://doi.org/10.1098/rstb.2005.1622
Froemke, R.C., 2015. Plasticity of Cortical Excitatory-Inhibitory Balance. Annual Review of Neuroscience 38, 195–219. https://doi.org/10.1146/annurev-neuro-071714-034002
Froemke, R.C., Martins, A.R.O., 2011. Spectrotemporal dynamics of auditory cortical synaptic receptive field plasticity. Hearing Research, Synaptic Plasticity 279, 149–161. https://doi.org/10.1016/j.heares.2011.03.005
Froemke, R.C., Merzenich, M.M., Schreiner, C.E., 2007. A synaptic memory trace for cortical receptive field plasticity. Nature 450, 425–429. https://doi.org/10.1038/nature06289
Fu, Y., Kaneko, M., Tang, Y., Alvarez-Buylla, A., Stryker, M.P., 2015. A cortical disinhibitory circuit for enhancing adult plasticity. eLife 4, e05558. https://doi.org/10.7554/eLife.05558
Fu, Y., Tucciarone, J.M., Espinosa, J.S., Sheng, N., Darcy, D.P., Nicoll, R.A., Huang, Z.J., Stryker, M.P., 2014. A cortical circuit for gain control by behavioral state. Cell 156, 1139–1152. https://doi.org/10.1016/j.cell.2014.01.050
Fuxe, K., Borroto-Escuela, D.O., Romero-Fernandez, W., Ciruela, F., Manger, P., Leo, G., Díaz-Cabiale, Z., Agnati, L.F., 2012. On the role of volume transmission and receptor-receptor interactions in social behaviour: focus on central catecholamine and oxytocin neurons. Brain Res 1476, 119–131. https://doi.org/10.1016/j.brainres.2012.01.062
Gabernet, L., Jadhav, S.P., Feldman, D.E., Carandini, M., Scanziani, M., 2005. Somatosensory Integration Controlled by Dynamic Thalamocortical Feed-Forward Inhibition. Neuron 48, 315–327. https://doi.org/10.1016/j.neuron.2005.09.022
Galarreta, M., Hestrin, S., 1999. A network of fast-spiking cells in the neocortex connected by electrical synapses. Nature 402, 72–75. https://doi.org/10.1038/47029
Garnock-Jones, K.P., Keating, G.M., 2009. Atomoxetine: a review of its use in attention-deficit hyperactivity disorder in children and adolescents. Paediatr Drugs 11, 203–226. https://doi.org/10.2165/00148581-200911030-00005
Garvert, M.M., Dolan, R.J., Behrens, T.E., 2017. A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. eLife 6, e17086. https://doi.org/10.7554/eLife.17086
Gasparovic, C., Song, T., Devier, D., Bockholt, H.J., Caprihan, A., Mullins, P.G., Posse, S., Jung, R., Morrison, L.A., 2006. Use of tissue water as a concentration reference for proton spectroscopic imaging. Magnetic resonance in medicine 55. https://doi.org/10.1002/mrm.20901
Gehlert, D.R., Gackenheimer, S.L., Robertson, D.W., 1993. Localization of rat brain binding sites for [3H]tomoxetine, an enantiomerically pure ligand for norepinephrine reuptake sites. Neuroscience Letters 157, 203–206. https://doi.org/10.1016/0304-3940(93)90737-6
Gentet, L.J., Kremer, Y., Taniguchi, H., Huang, Z.J., Staiger, J.F., Petersen, C.C.H., 2012. Unique functional properties of somatostatin-expressing GABAergic neurons in mouse barrel cortex. Nat Neurosci 15, 607–612. https://doi.org/10.1038/nn.3051
Gilbert, P.E., Kesner, R.P., Lee, I., 2001. Dissociating hippocampal subregions: double dissociation between dentate gyrus and CA1. Hippocampus 11, 626–636. https://doi.org/10.1002/hipo.1077
Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M., Cohen, J.D., 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. Cogn Affect Behav Neurosci 10, 252–269. https://doi.org/10.3758/CABN.10.2.252
Gjedde, A., Marrett, S., Vafaee, M., 2002. Oxidative and Nonoxidative Metabolism of Excited Neurons and Astrocytes. J Cereb Blood Flow Metab 22, 1–14. https://doi.org/10.1097/00004647-200201000-00001

Godlewska, B.R., Clare, S., Cowen, P.J., Emir, U.E., 2017. Ultra-High-Field Magnetic Resonance Spectroscopy in Psychiatry. Frontiers in Psychiatry 8.
Gogolla, N., Leblanc, J.J., Quast, K.B., Südhof, T.C., Fagiolini, M., Hensch, T.K., 2009. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. J Neurodev Disord 1, 172–181. https://doi.org/10.1007/s11689-009-9023-x
Gold, A.E., Kesner, R.P., 2005. The role of the CA3 subregion of the dorsal hippocampus in spatial pattern completion in the rat. Hippocampus 15, 808–814. https://doi.org/10.1002/hipo.20103
Goode, T.D., Tanaka, K.Z., Sahay, A., McHugh, T.J., 2020. An Integrated Index: Engrams, Place Cells, and Hippocampal Memory. Neuron 107, 805–820. https://doi.org/10.1016/j.neuron.2020.07.011
Govindaraju, V., Young, K., Maudsley, A.A., 2000. Proton NMR chemical shifts and coupling constants for brain metabolites. NMR in Biomedicine 13, 129–153. https://doi.org/10.1002/1099-1492(200005)13:3
Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn Sci 10, 14–23. https://doi.org/10.1016/j.tics.2005.11.006
Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. Vision Research 41, 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6
Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R., 1999. Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. Neuron 24, 187–203. https://doi.org/10.1016/S0896-6273(00)80832-6
Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. Acta Psychologica, Beyond the decade of the brain: Towards functional neuronanatomy of the mind 107, 293–321. https://doi.org/10.1016/S0001-6918(01)00019-1
Gruetter, R., Tkác, I., 2000. Field mapping without reference scan using asymmetric echo-planar techniques. Magn Reson Med 43, 319–323. https://doi.org/10.1002/(sici)1522-2594(200002)43:2<319::aid-mrm22>3.0.co;2-1
Gu, Z., Yakel, J.L., 2011. Timing-dependent septal cholinergic induction of dynamic hippocampal synaptic plasticity. Neuron 71, 155–165. https://doi.org/10.1016/j.neuron.2011.04.026
Gussew, A., Rzanny, R., Erdtel, M., Scholle, H.C., Kaiser, W.A., Mentzel, H.J., Reichenbach, J.R., 2010. Time-resolved functional 1H MR spectroscopic detection of glutamate concentration changes in the brain during acute heat pain stimulation. NeuroImage 49, 1895–1902. https://doi.org/10.1016/j.neuroimage.2009.09.007
Haase, A., Frahm, J., Hanicke, W., Matthaei, D., 1985. 1H NMR chemical shift selective (CHESS) imaging. Phys. Med. Biol. 30, 341–344. https://doi.org/10.1088/0031-9155/30/4/008
Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., Moser, E.I., 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806. https://doi.org/10.1038/nature03721
Haider, B., Duque, A., Hasenstaub, A.R., McCormick, D.A., 2006. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. J. Neurosci. 26, 4535–4545. https://doi.org/10.1523/JNEUROSCI.5297-05.2006
Hasselmo, M.E., 2006. The Role of Acetylcholine in Learning and Memory. Curr Opin Neurobiol 16, 710–715. https://doi.org/10.1016/j.conb.2006.09.002

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293, 2425–2430. https://doi.org/10.1126/science.1063736
Hebb, D.O., 1949. The Organization of Behavior: A Neuropsychological Theory. Psychology Press, New York. https://doi.org/10.4324/9781410612403
Heckers, S., Konradi, C., 2015. GABAergic mechanisms of hippocampal hyperactivity in schizophrenia. Schizophr Res 167, 4–11. https://doi.org/10.1016/j.schres.2014.09.041
Heldt, S.A., Ressler, K.J., 2007. Training-induced changes in the expression of GABAA-associated genes in the amygdala after the acquisition and extinction of Pavlovian fear. Eur J Neurosci 26, 3631–3644. https://doi.org/10.1111/j.1460-9568.2007.05970.x
Hennequin, G., Agnes, E.J., Vogels, T.P., 2017. Inhibitory Plasticity: Balance, Control, and Codependence. Annu Rev Neurosci 40, 557–579. https://doi.org/10.1146/annurev-neuro-072116-031005
Hess, A.T., Tisdall, M.D., Andronesi, O.C., Meintjes, E.M., Van Der Kouwe, A.J.W., 2011. Real-Time Motion and B0 Corrected Single Voxel Spectroscopy Using Volumetric Navigators. Magnetic Resonance in Medicine 66, 314–323. https://doi.org/10.1002/mrm.22805
Hiramatsu, C., Goda, N., Komatsu, H., 2011. Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. NeuroImage 57, 482–494. https://doi.org/10.1016/j.neuroimage.2011.04.056
Ho, J., Tumkaya, T., Aryal, S., Choi, H., Claridge-Chang, A., 2019. Moving beyond P values: data analysis with estimation graphics. Nat Methods 16, 565–566. https://doi.org/10.1038/s41592-019-0470-3
Hoffman, E.J., Warren, E.W., 1993. Flumazenil: a benzodiazepine antagonist. Clin Pharm 12, 641–56; quiz 699–701.
Hong, D., Rankouhi, S.R., Thielen, J.-W., Asten, J.J.A. van, Norris, D.G., 2019. A comparison of sLASER and MEGA-sLASER using simultaneous interleaved acquisition for measuring GABA in the human brain at 7T. PLOS ONE 14, e0223702. https://doi.org/10.1371/journal.pone.0223702
Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79, 2554–2558.
Horner, A.J., Bisby, J.A., Bush, D., Lin, W.-J., Burgess, N., 2015. Evidence for holistic episodic recollection via hippocampal pattern completion. Nature Communications 6, 1–11. https://doi.org/10.1038/ncomms8462
Howard, M.W., Rizzuto, D.S., Caplan, J.B., Madsen, J.R., Lisman, J., Aschenbrenner-Scheibe, R., Schulze-Bonhage, A., Kahana, M.J., 2003. Gamma Oscillations Correlate with Working Memory Load in Humans. Cerebral Cortex 13, 1369–1374. https://doi.org/10.1093/cercor/bhg084
Howarth, C., Mishra, A., Hall, C.N., 2021. More than just summed neuronal activity: how multiple cell types shape the BOLD response. Philosophical Transactions of the Royal Society B: Biological Sciences 376, 20190630. https://doi.org/10.1098/rstb.2019.0630
Hsieh, L.-T., Gruber, M.J., Jenkins, L.J., Ranganath, C., 2014. Hippocampal activity patterns carry information about objects in temporal context. Neuron 81, 1165–1178. https://doi.org/10.1016/j.neuron.2014.01.015
Hu, H., Gan, J., Jonas, P., 2014. Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function. Science 345, 1255263. https://doi.org/10.1126/science.1255263
Huffman, D.J., Stark, C.E.L., 2014. Multivariate pattern analysis of the human medial temporal lobe revealed representationally categorical cortex and representationally agnostic hippocampus. Hippocampus 24, 1394–1403. https://doi.org/10.1002/hipo.22321

Hummel, F.C., Cohen, L.G., 2006. Non-invasive brain stimulation: a new strategy to improve neurorehabilitation after stroke? The Lancet Neurology 5, 708–712. https://doi.org/10.1016/S1474-4422(06)70525-7
Insausti, R., Herrero, M.T., Witter, M.P., 1997. Entorhinal cortex of the rat: Cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. Hippocampus 7, 146–183. https://doi.org/10.1002/(SICI)1098-1063(1997)7:2<146::AID-HIPO4>3.0.CO;2-L
Ip, I.B., Berrington, A., Hess, A.T., Parker, A.J., Emir, U.E., Bridge, H., 2017. Combined fMRI-MRS acquires simultaneous glutamate and BOLD-fMRI signals in the human brain. NeuroImage 155, 113–119. https://doi.org/10.1016/j.neuroimage.2017.04.030
Ip, I.B., Emir, U.E., Parker, A.J., Campbell, J., Bridge, H., 2019. Comparison of Neurochemical and BOLD Signal Contrast Response Functions in the Human Visual Cortex. J. Neurosci. 39, 7968–7975. https://doi.org/10.1523/JNEUROSCI.3021-18.2019
Izquierdo, I., Medina, J.H., Bianchin, M., Walz, R., Zanatta, M.S., Da Silva, R.C., Bueno e Silva, M., Ruschel, A.C., Paczko, N., 1993. Memory processing by the limbic system: role of specific neurotransmitter systems. Behav Brain Res 58, 91–98. https://doi.org/10.1016/0166-4328(93)90093-6
Jacobs, J., Miller, J., Lee, S.A., Coffey, T., Watrous, A.J., Sperling, M.R., Sharan, A., Worrell, G., Berry, B., Lega, B., Jobst, B.C., Davis, K., Gross, R.E., Sheth, S.A., Ezzyat, Y., Das, S.R., Stein, J., Gorniak, R., Kahana, M.J., Rizzuto, D.S., 2016. Direct Electrical Stimulation of the Human Entorhinal Region and Hippocampus Impairs Memory. Neuron 92, 983–990. https://doi.org/10.1016/j.neuron.2016.10.062
Jacobs, K.M., Donoghue, J.P., 1991. Reshaping the cortical motor map by unmasking latent intracortical connections. Science 251, 944–947. https://doi.org/10.1126/science.2000496
Jacobson, L., Koslowsky, M., Lavidor, M., 2012. tDCS polarity effects in motor and cognitive domains: a meta-analytical review. Exp Brain Res 216, 1–10. https://doi.org/10.1007/s00221-011-2891-9
Jelen, L.A., King, S., Mullins, P.G., Stone, J.M., 2018. Beyond static measures: A review of functional magnetic resonance spectroscopy and its potential to investigate dynamic glutamatergic abnormalities in schizophrenia. Journal of Psychopharmacology 32, 497–508.
Ji, D., Wilson, M.A., 2007. Coordinated memory replay in the visual cortex and hippocampus during sleep. Nat Neurosci 10, 100–107. https://doi.org/10.1038/nn1825
Johnson, K.A., Conn, P.J., Niswender, C.M., 2009. Glutamate receptors as therapeutic targets for Parkinson's disease. CNS Neurol Disord Drug Targets 8, 475–491.
Joshi, S., Li, Y., Kalwani, R.M., Gold, J.I., 2016. Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex. Neuron 89, 221–234. https://doi.org/10.1016/j.neuron.2015.11.028
Kalm, K., Davis, M.H., Norris, D., 2013. Individual Sequence Representations in the Medial Temporal Lobe. Journal of Cognitive Neuroscience 25, 1111–1121. https://doi.org/10.1162/jocn_a_00378
Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nat Neurosci 8, 679–685. https://doi.org/10.1038/nn1444
Karlsson, M.P., Frank, L.M., 2009. Awake replay of remote experiences in the hippocampus. Nat Neurosci 12, 913–918. https://doi.org/10.1038/nn.2344
Karnani, M.M., Jackson, J., Ayzenshtat, I., Hamzehei Sichani, A., Manoocheri, K., Kim, S., Yuste, R., 2016a. Opening Holes in the Blanket of Inhibition: Localized Lateral Disinhibition by VIP Interneurons. J Neurosci 36, 3471–3480. https://doi.org/10.1523/JNEUROSCI.3646-15.2016

Karnani, M.M., Jackson, J., Ayzenshtat, I., Tucciarone, J., Manoocheri, K., Snider, W.G., Yuste, R., 2016b. Cooperative Subnetworks of Molecularly Similar Interneurons in Mouse Neocortex. Neuron 90, 86–100. https://doi.org/10.1016/j.neuron.2016.02.037
Katragadda, S., Schubiner, H., 2007. ADHD in children, adolescents, and adults. Prim Care 34, 317–341; abstract viii. https://doi.org/10.1016/j.pop.2007.04.012
Kauppinen, R.A., Pirttilä, T.R., Auriola, S.O., Williams, S.R., 1994. Compartmentation of cerebral glutamate in situ as detected by 1H/13C n.m.r. Biochem J 298, 121–127.
Kauppinen, R.A., Williams, S.R., 1991. Nondestructive Detection of Glutamate by 1H Nuclear Magnetic Resonance Spectroscopy in Cortical Brain Slices from the Guinea Pig: Evidence for Changes in Detectability During Severe Anoxic Insults. Journal of Neurochemistry 57, 1136–1144. https://doi.org/10.1111/j.1471-4159.1991.tb08271.x
Kawaguchi, Y., Katsumaru, H., Kosaka, T., Heizmann, C.W., Hama, K., 1987. Fast spiking cells in rat hippocampus (CA1 region) contain the calcium-binding protein parvalbumin. Brain Research 416, 369–374. https://doi.org/10.1016/0006-8993(87)90921-8
Kawaguchi, Y., Kubota, Y., 1997. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cerebral Cortex 7, 476–486. https://doi.org/10.1093/cercor/7.6.476
Kebschull, J.M., Garcia da Silva, P., Reid, A.P., Peikon, I.D., Albeanu, D.F., Zador, A.M., 2016. High-Throughput Mapping of Single-Neuron Projections by Sequencing of Barcoded RNA. Neuron 91, 975–987. https://doi.org/10.1016/j.neuron.2016.07.036
Khan, A.G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T.D., Hofer, S.B., 2018. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. Nat Neurosci 21, 851–859. https://doi.org/10.1038/s41593-018-0143-z
Kim, J.-H., Jung, A.-H., Jeong, D., Choi, I., Kim, K., Shin, S., Kim, S.J., Lee, S.-H., 2016. Selectivity of Neuromodulatory Projections from the Basal Forebrain and Locus Ceruleus to Primary Sensory Cortices. J. Neurosci. 36, 5314–5327. https://doi.org/10.1523/JNEUROSCI.4333-15.2016
Kim, J.J., Fanselow, M.S., 1992. Modality-Specific Retrograde Amnesia of Fear. Science 256, 675–677. https://doi.org/10.1126/science.1585183
Kim, S., Stephenson, M.C., Morris, P.G., Jackson, S.R., 2014. tDCS-induced alterations in GABA concentration within primary motor cortex predict motor learning and motor memory: A 7T magnetic resonance spectroscopy study. NeuroImage 99, 237–243. https://doi.org/10.1016/j.neuroimage.2014.05.070
Kirkwood, A., Dudek, S.M., Gold, J.T., Aizenman, C.D., Bear, M.F., 1993. Common Forms of Synaptic Plasticity in the Hippocampus and Neocortex in Vitro. Science 260, 1518–1521. https://doi.org/10.1126/science.8502997
Klausberger, T., Somogyi, P., 2008. Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations. Science 321, 53–57. https://doi.org/10.1126/science.1149381
Koda, K., Ago, Y., Cong, Y., Kita, Y., Takuma, K., Matsuda, T., 2010. Effects of acute and chronic administration of atomoxetine and methylphenidate on extracellular levels of noradrenaline, dopamine and serotonin in the prefrontal cortex and striatum of mice. Journal of Neurochemistry 114, 259–270. https://doi.org/10.1111/j.1471-4159.2010.06750.x
Koechlin, E., Ody, C., Kouneiher, F., 2003. The Architecture of Cognitive Control in the Human Prefrontal Cortex. Science 302, 1181–1185. https://doi.org/10.1126/science.1088545
Kolasinski, J., Hinson, E.L., Zand, A.P.D., Rizov, A., Emir, U.E., Stagg, C.J., 2019. The dynamics of cortical GABA in human motor learning. The Journal of Physiology 597, 271–282. https://doi.org/10.1113/JP276626

Komorowski, R.W., Manns, J.R., Eichenbaum, H., 2009. Robust conjunctive item-place coding by hippocampal neurons parallels learning what happens where. J Neurosci 29, 9918–9929. https://doi.org/10.1523/JNEUROSCI.1378-09.2009
Koolschijn, R.S., Emir, U.E., Pantelides, A.C., Nili, H., Behrens, T.E.J., Barron, H.C., 2019. The Hippocampus and Neocortical Inhibitory Engrams Protect against Memory Interference. Neuron 101, 528–541. https://doi.org/10.1016/j.neuron.2018.11.042
Koolschijn, R.S.*, Shpektor, A.*, Clarke, W.T., Ip, I.B., Dupret, D., Emir, U.E., Barron, H.C., 2021. Memory recall involves a transient break in excitatory-inhibitory balance. eLife 10, e70071. https://doi.org/10.7554/eLife.70071
Kosten, T.R., Mason, J.W., Giller, E.L., Ostroff, R.B., Harkness, L., 1987. Sustained urinary norepinephrine and epinephrine elevation in post-traumatic stress disorder. Psychoneuroendocrinology 12, 13–20. https://doi.org/10.1016/0306-4530(87)90017-5
Koster, R., Chadwick, M.J., Chen, Y., Berron, D., Banino, A., Düzel, E., Hassabis, D., Kumaran, D., 2018. Big-Loop Recurrence within the Hippocampal System Supports Integration of Information across Episodes. Neuron 99, 1342-1354.e6. https://doi.org/10.1016/j.neuron.2018.08.009
Krabbe, S., Paradiso, E., d'Aquin, S., Bitterman, Y., Courtin, J., Xu, C., Yonehara, K., Markovic, M., Müller, C., Eichlisberger, T., Gründemann, J., Ferraguti, F., Lüthi, A., 2019. Adaptive disinhibitory gating by VIP interneurons permits associative learning. Nature Neuroscience 22, 1834–1843. https://doi.org/10.1038/s41593-019-0508-y
Krach, L.E., 2001. Pharmacotherapy of Spasticity: Oral Medications and Intrathecal Baclofen. J Child Neurol 16, 31–36. https://doi.org/10.1177/088307380101600106
Kreis, R., 2004. Issues of spectral quality in clinical 1H-magnetic resonance spectroscopy and a gallery of artifacts. NMR in Biomedicine 17, 361–381. https://doi.org/10.1002/nbm.891
Krekelberg, B., Boynton, G.M., van Wezel, R.J.A., 2006. Adaptation: from single cells to BOLD signals. Trends Neurosci 29, 250–256. https://doi.org/10.1016/j.tins.2006.02.008
Kriegeskorte, N., Mur, M., Bandettini, P., 2008a. Representational similarity analysis - connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience 2. https://doi.org/10.3389/neuro.06.004.2008
Kriegeskorte, N., Mur, M., Ruff, D.A., Kiani, R., Bodurka, J., Esteky, H., Tanaka, K., Bandettini, P.A., 2008b. Matching Categorical Object Representations in Inferior Temporal Cortex of Man and Monkey. Neuron 60, 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043
Kroon, J.A.V., Carobrez, A.P., 2009. Olfactory fear conditioning paradigm in rats: effects of midazolam, propranolol or scopolamine. Neurobiol Learn Mem 91, 32–40. https://doi.org/10.1016/j.nlm.2008.10.007
Krupic, J., Bauza, M., Burton, S., Barry, C., O'Keefe, J., 2015. Grid cell symmetry is shaped by environmental geometry. Nature 518, 232–235. https://doi.org/10.1038/nature14153
Kuchibhotla, K.V., Gill, J.V., Lindsay, G.W., Papadoyannis, E.S., Field, R.E., Sten, T.A.H., Miller, K.D., Froemke, R.C., 2017. Parallel processing by cortical inhibition enables context-dependent behavior. Nature Neuroscience 20, 62–71. https://doi.org/10.1038/nn.4436
Kuhlman, S.J., Olivas, N.D., Tring, E., Ikrar, T., Xu, X., Trachtenberg, J.T., 2013. A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex. Nature 501, 543–546. https://doi.org/10.1038/nature12485
Kuo, S.P., Trussell, L.O., 2011. Spontaneous Spiking and Synaptic Depression Underlie Noradrenergic Control of Feed-Forward Inhibition. Neuron 71, 306–318. https://doi.org/10.1016/j.neuron.2011.05.039
Lally, N., Mullins, P.G., Roberts, M.V., Price, D., Gruber, T., Haenschel, C., 2014. Glutamatergic correlates of gamma-band oscillatory activity during cognition: A concurrent ER-MRS

and EEG study. NeuroImage, New Horizons for Neural Oscillations 85, 823–833. https://doi.org/10.1016/j.neuroimage.2013.07.049
Lea-Carnall, C.A., El-Deredy, W., Williams, S.R., Stagg, C.J., Trujillo-Barreto, N.J., 2021. From synaptic activity to human in vivo quantification of neurotransmitter dynamics: a neural modelling approach. https://doi.org/10.1101/2021.03.11.434540
Lee, I., Kesner, R.P., 2004. Encoding versus retrieval of spatial memory: double dissociation between the dentate gyrus and the perforant path inputs into CA3 in the dorsal hippocampus. Hippocampus 14, 66–76. https://doi.org/10.1002/hipo.10167
Lee, L., Boorman, L., Glendenning, E., Christmas, C., Sharp, P., Redgrave, P., Shabir, O., Bracci, E., Berwick, J., Howarth, C., 2020. Key Aspects of Neurovascular Control Mediated by Specific Populations of Inhibitory Cortical Interneurons. Cerebral Cortex 30, 2452–2464. https://doi.org/10.1093/cercor/bhz251
Lee, S., Kruglikov, I., Huang, Z.J., Fishell, G., Rudy, B., 2013. A disinhibitory circuit mediates motor integration in the somatosensory cortex. Nature Neuroscience 16, 1662–1670. https://doi.org/10.1038/nn.3544
Letzkus, J.J., Wolff, S.B.E., Lüthi, A., 2015. Disinhibition, a Circuit Mechanism for Associative Learning and Memory. Neuron 88, 264–276. https://doi.org/10.1016/j.neuron.2015.09.024
Letzkus, J.J., Wolff, S.B.E., Meyer, E.M.M., Tovote, P., Courtin, J., Herry, C., Lüthi, A., 2011. A disinhibitory microcircuit for associative fear learning in the auditory cortex. Nature 480, 331–335. https://doi.org/10.1038/nature10674
Leutgeb, J.K., Leutgeb, S., Moser, M.-B., Moser, E.I., 2007. Pattern separation in the dentate gyrus and CA3 of the hippocampus. Science 315, 961–966. https://doi.org/10.1126/science.1135801
Leutgeb, S., Leutgeb, J.K., Moser, M.-B., Moser, E.I., 2005. Place cells, spatial maps and the population code for memory. Current Opinion in Neurobiology 15, 738–746. https://doi.org/10.1016/j.conb.2005.10.002
Li, L., Miller, E.K., Desimone, R., 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol 69, 1918–1929. https://doi.org/10.1152/jn.1993.69.6.1918
Lin, Andronesi, O., Bogner, W., Choi, I.-Y., Coello, E., Cudalbu, C., Juchem, C., Kemp, G.J., Kreis, R., Krššák, M., Lee, P., Maudsley, A.A., Meyerspeer, M., Mlynarik, V., Near, J., Öz, G., Peek, A.L., Puts, N.A., Ratai, E.-M., Tkáč, I., Mullins, P.G., 2021. Minimum Reporting Standards for in vivo Magnetic Resonance Spectroscopy (MRSinMRS): Experts' consensus recommendations. NMR in Biomedicine 34, e4484. https://doi.org/10.1002/nbm.4484
Lin, Y., Stephenson, M.C., Xin, L., Napolitano, A., Morris, P.G., 2012. Investigating the Metabolic Changes due to Visual Stimulation using Functional Proton Magnetic Resonance Spectroscopy at 7 T. J Cereb Blood Flow Metab 32, 1484–1495. https://doi.org/10.1038/jcbfm.2012.33
Lindner, M., Bell, T., Iqbal, S., Mullins, P.G., Christakou, A., 2017. In vivo functional neurochemistry of human cortical cholinergic function during visuospatial attention. PLoS One 12, e0171338. https://doi.org/10.1371/journal.pone.0171338
Linkerhand, M., Gros, C., 2013. Generating functionals for autonomous latching dynamics in attractor relict networks. Sci Rep 3, 2042. https://doi.org/10.1038/srep02042
Lisman, J.E., Coyle, J.T., Green, R.W., Javitt, D.C., Benes, F.M., Heckers, S., Grace, A.A., 2008. Circuit-based framework for understanding neurotransmitter and risk gene interactions in schizophrenia. Trends Neurosci 31, 234–242. https://doi.org/10.1016/j.tins.2008.02.005

Litwin-Kumar, A., Doiron, B., 2014. Formation and maintenance of neuronal assemblies through synaptic plasticity. Nat Commun 5, 5319. https://doi.org/10.1038/ncomms6319
Logothetis, N.K., 2003. The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal. J. Neurosci. 23, 3963–3971. https://doi.org/10.1523/jneurosci.23-10-03963.2003
Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal. Nature 412, 150–157. https://doi.org/10.1038/35084005
Losonczy, A., Zemelman, B.V., Vaziri, A., Magee, J.C., 2010. Network mechanisms of theta related neuronal activity in hippocampal CA1 pyramidal neurons. Nat Neurosci 13, 967–972. https://doi.org/10.1038/nn.2597
Louie, K., Wilson, M.A., 2001. Temporally Structured Replay of Awake Hippocampal Ensemble Activity during Rapid Eye Movement Sleep. Neuron 29, 145–156. https://doi.org/10.1016/S0896-6273(01)00186-6
Lovett-Barron, M., Kaifosh, P., Kheirbek, M.A., Danielson, N., Zaremba, J.D., Reardon, T.R., Turi, G.F., Hen, R., Zemelman, B.V., Losonczy, A., 2014. Dendritic inhibition in the hippocampus supports fear learning. Science 343, 857–863. https://doi.org/10.1126/science.1247485
Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., Miller, E.K., 2016. Gamma and Beta Bursts Underlie Working Memory. Neuron 90, 152–164. https://doi.org/10.1016/j.neuron.2016.02.028
Lunghi, C., Berchicci, M., Morrone, M.C., Russo, F.D., 2015. Short-term monocular deprivation alters early components of visual evoked potentials. The Journal of Physiology 593, 4361–4372. https://doi.org/10.1113/JP270950
Ma, Y., Hu, H., Berrebi, A.S., Mathers, P.H., Agmon, A., 2006. Distinct subtypes of somatostatin-containing neocortical interneurons revealed in transgenic mice. J Neurosci 26, 5069–5082. https://doi.org/10.1523/jneurosci.0661-06.2006
MacDonald, C.J., Lepage, K.Q., Eden, U.T., Eichenbaum, H., 2011. Hippocampal "Time Cells" Bridge the Gap in Memory for Discontiguous Events. Neuron 71, 737–749. https://doi.org/10.1016/j.neuron.2011.07.012
Magistretti, P.J., Allaman, I., 2015. A cellular perspective on brain energy metabolism and functional imaging. Neuron 86, 883–901. https://doi.org/10.1016/j.neuron.2015.03.035
Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F.Q., Leopold, D.A., 2008. Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. Nat Neurosci 11, 1193–1200. https://doi.org/10.1038/nn.2173
Makino, H., Komiyama, T., 2015. Learning enhances the relative impact of top-down processing in the visual cortex. Nat Neurosci 18, 1116–1122. https://doi.org/10.1038/nn.4061
Maksymetz, J., Moran, S.P., Conn, P.J., 2017. Targeting metabotropic glutamate receptors for novel treatments of schizophrenia. Molecular Brain 10, 15. https://doi.org/10.1186/s13041-017-0293-z
Malenka, R.C., Nicoll, and R.A., 1999. Long-Term Potentiation--A Decade of Progress? Science 285, 1870–1874. https://doi.org/10.1126/science.285.5435.1870
Malik, R., Li, Y., Schamiloglu, S., Sohal, V.S., 2022. Top-down control of hippocampal signal-to-noise by prefrontal long-range inhibition. Cell 185, 1602-1617.e17. https://doi.org/10.1016/j.cell.2022.04.001
Mangia, S., Tkáč, I., Gruetter, R., Van De Moortele, P.-F., Giove, F., Maraviglia, B., Uğurbil, K., 2006. Sensitivity of single-voxel 1H-MRS in investigating the metabolism of the activated human visual cortex at 7 T. Magnetic Resonance Imaging 24, 343–348. https://doi.org/10.1016/j.mri.2005.12.023

Manunta, Y., Edeline, J.-M., 2004. Noradrenergic induction of selective plasticity in the frequency tuning of auditory cortex neurons. J Neurophysiol 92, 1445–1463. https://doi.org/10.1152/jn.00079.2004
Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., Wu, C., 2004. Interneurons of the neocortical inhibitory system. Nat Rev Neurosci 5, 793–807. https://doi.org/10.1038/nrn1519
Marr, D., 1971. Simple memory: a theory for archicortex. Philos Trans R Soc Lond B Biol Sci 262, 23–81. https://doi.org/10.1098/rstb.1971.0078
Martins, A.R.O., Froemke, R.C., 2015. Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. Nat Neurosci 18, 1483–1492. https://doi.org/10.1038/nn.4090
McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102, 419–457. https://doi.org/10.1037/0033-295X.102.3.419
McCloskey, M., Cohen, N.J., 1989. Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem, in: Bower, G.H. (Ed.), Psychology of Learning and Motivation. Academic Press, pp. 109–165. https://doi.org/10.1016/S0079-7421(08)60536-8
McKenzie, S., Frank, A.J., Kinsky, N.R., Porter, B., Rivière, P.D., Eichenbaum, H., 2014. Hippocampal representation of related and opposing memories develop within distinct, hierarchically-organized neural schemas. Neuron 83, 202–215. https://doi.org/10.1016/j.neuron.2014.05.019
McNamara, C.G., Tejero-Cantero, Á., Trouche, S., Campo-Urriza, N., Dupret, D., 2014. Dopaminergic neurons promote hippocampal reactivation and spatial memory persistence. Nat Neurosci 17, 1658–1660. https://doi.org/10.1038/nn.3843
McNamara, T.P., 1986. Mental representations of spatial relations. Cognitive Psychology 18, 87–121. https://doi.org/10.1016/0010-0285(86)90016-2
Mechelli, A., Price, C.J., Henson, R.N.A., Friston, K.J., 2003. Estimating efficiency a priori: a comparison of blocked and randomized designs. Neuroimage 18, 798–805. https://doi.org/10.1016/s1053-8119(02)00040-x
Megemont, M., McBurney-Lin, J., Yang, H., 2022. Pupil diameter is not an accurate real-time readout of locus coeruleus activity. eLife 11, e70510. https://doi.org/10.7554/eLife.70510
Mekle, R., Kühn, S., Pfeiffer, H., Aydin, S., Schubert, F., Ittermann, B., 2017. Detection of metabolite changes in response to a varying visual stimulation paradigm using short-TE 1H MRS at 7 T. NMR in Biomedicine 30, e3672. https://doi.org/10.1002/nbm.3672
Mescher, M., Merkle, H., Kirsch, J., Garwood, M., Gruetter, R., 1998. Simultaneous in vivo spectral editing and water suppression. NMR in Biomedicine 11, 266–272. https://doi.org/10.1002/(SICI)1099-1492(199810)11:6<266::AID-NBM530>3.0.CO;2-J
Metherate, R., Cox, C.L., Ashe, J.H., 1992. Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine. J. Neurosci. 12, 4701–4711. https://doi.org/10.1523/jneurosci.12-12-04701.1992
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu Rev Neurosci 24, 167–202. https://doi.org/10.1146/annurev.neuro.24.1.167
Miller, E.K., Li, L., Desimone, R., 1991. A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. Science 254, 1377–1379. https://doi.org/10.1126/science.1962197
Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.M., Daw, N.D., Gershman, S.J., 2017. The successor representation in human reinforcement learning. Nat Hum Behav 1, 680–692. https://doi.org/10.1038/s41562-017-0180-8

Mondrup, K., Pedersen, E., 1984. The clinical effect of the GABA-agonist, progabide, on spasticity. Acta Neurologica Scandinavica 69, 200–206. https://doi.org/10.1111/j.1600-0404.1984.tb07802.x
Moore, C.I., Carlen, M., Knoblich, U., Cardin, J.A., 2010. Neocortical Interneurons: From Diversity, Strength. Cell 142, 184–188. https://doi.org/10.1016/j.cell.2010.07.005
Moradi, F., Buxton, R.B., 2013. Adaptation of cerebral oxygen metabolism and blood flow and modulation of neurovascular coupling with prolonged stimulation in human visual cortex. Neuroimage 82, 182–189. https://doi.org/10.1016/j.neuroimage.2013.05.110
Morris, R.G.M., Frey, U., 1997. Hippocampal synaptic plasticity: role in spatial learning or the automatic recording of attended experience? Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 352, 1489–1503. https://doi.org/10.1098/rstb.1997.0136
Morton, N.W., Sherrill, K.R., Preston, A.R., 2017. Memory integration constructs maps of space, time, and concepts. Current Opinion in Behavioral Sciences, Memory in time and space 17, 161–168. https://doi.org/10.1016/j.cobeha.2017.08.007
Mosha, N., Robertson, E.M., 2016. Unstable Memories Create a High-Level Representation that Enables Learning Transfer. Current Biology 26, 100–105. https://doi.org/10.1016/j.cub.2015.11.035
Mullins, P.G., 2018. Towards a theory of functional magnetic resonance spectroscopy (fMRS): A meta-analysis and discussion of using MRS to measure changes in neurotransmitters in real time. Scandinavian Journal of Psychology 59, 91–103. https://doi.org/10.1111/sjop.12411
Mullins, P.G., McGonigle, D.J., O'Gorman, R.L., Puts, N.A.J., Vidyasagar, R., Evans, C.J., Edden, R.A.E., 2014. Current practice in the use of MEGA-PRESS spectroscopy for the detection of GABA. NeuroImage 86, 43–52. https://doi.org/10.1016/j.neuroimage.2012.12.004
Murayama, M., Pérez-Garci, E., Nevian, T., Bock, T., Senn, W., Larkum, M.E., 2009. Dendritic encoding of sensory stimuli controlled by deep cortical interneurons. Nature 457, 1137–1141. https://doi.org/10.1038/nature07663
Murchison, C.F., Zhang, X.-Y., Zhang, W.-P., Ouyang, M., Lee, A., Thomas, S.A., 2004. A distinct role for norepinephrine in memory retrieval. Cell 117, 131–143. https://doi.org/10.1016/s0092-8674(04)00259-4
Myers, J.F., Nutt, D.J., Lingford-Hughes, A.R., 2016. γ-aminobutyric acid as a metabolite: Interpreting magnetic resonance spectroscopy experiments. J Psychopharmacol 30, 422–427. https://doi.org/10.1177/0269881116639298
Nabavi, S., Fox, R., Proulx, C.D., Lin, J.Y., Tsien, R.Y., Malinow, R., 2014. Engineering a memory with LTD and LTP. Nature 511, 348–352. https://doi.org/10.1038/nature13294
Nádasdy, Z., Hirase, H., Czurkó, A., Csicsvari, J., Buzsáki, G., 1999. Replay and Time Compression of Recurring Spike Sequences in the Hippocampus. J. Neurosci. 19, 9497–9507. https://doi.org/10.1523/jneurosci.19-21-09497.1999
Nadel, L., 2008. The hippocampus and context revisited, in: Hippocampal Place Fields: Relevance to Learning and Memory. Oxford University Press, New York, NY, US, pp. 3–15. https://doi.org/10.1093/acprof:oso/9780195323245.003.0002
Nadel, L., Moscovitch, M., 1997. Memory consolidation, retrograde amnesia and the hippocampal complex. Current Opinion in Neurobiology 7, 217–227. https://doi.org/10.1016/S0959-4388(97)80010-4
Nakazawa, K., Quirk, M.C., Chitwood, R.A., Watanabe, M., Yeckel, M.F., Sun, L.D., Kato, A., Carr, C.A., Johnston, D., Wilson, M.A., Tonegawa, S., 2002. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. Science 297, 211–218. https://doi.org/10.1126/science.1071795

Nassar, M.R., Rumsey, K.M., Wilson, R.C., Parikh, K., Heasly, B., Gold, J.I., 2012. Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15, 1040–1046. https://doi.org/10.1038/nn.3130
Natt, O., Bezkorovaynyy, V., Michaelis, T., Frahm, J., 2005. Use of phased array coils for a determination of absolute metabolite concentrations. Magn Reson Med 53, 3–8. https://doi.org/10.1002/mrm.20337
Navawongse, R., Eichenbaum, H., 2013. Distinct Pathways for Rule-Based Retrieval and Spatial Mapping of Memory Representations in Hippocampal Neurons. J. Neurosci. 33, 1002–1013. https://doi.org/10.1523/jneurosci.3891-12.2013
Near, J., Andersson, J., Maron, E., Mekle, R., Gruetter, R., Cowen, P., Jezzard, P., 2013. Unedited in vivo detection and quantification of γ-aminobutyric acid in the occipital cortex using short-TE MRS at 3 T. NMR Biomed 26, 1353–1362. https://doi.org/10.1002/nbm.2960
Nelson, S.B., Valakh, V., 2015. Excitatory/Inhibitory balance and circuit homeostasis in Autism Spectrum Disorders. Neuron 87, 684–698. https://doi.org/10.1016/j.neuron.2015.07.033
Nicola, W., Clopath, C., 2019. A diversity of interneurons and Hebbian plasticity facilitate rapid compressible learning in the hippocampus. Nat Neurosci 22, 1168–1181. https://doi.org/10.1038/s41593-019-0415-2
Nikolova, S., Stark, S.M., Stark, C.E.L., 2017. 3T hippocampal glutamate-glutamine complex reflects verbal memory decline in aging. Neurobiology of Aging 54, 103–111. https://doi.org/10.1016/j.neurobiolaging.2017.01.026
Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., 2014. A Toolbox for Representational Similarity Analysis. PLOS Computational Biology 10, e1003553. https://doi.org/10.1371/journal.pcbi.1003553
Nishitani, N., 2003. Dynamics of cognitive processing in the human hippocampus by neuromagnetic and neurochemical assessments. NeuroImage 20, 561–571. https://doi.org/10.1016/S1053-8119(03)00280-5
Nishiyama, M., Hong, K., Mikoshiba, K., Poo, M., Kato, K., 2000. Calcium stores regulate the polarity and input specificity of synaptic modification. Nature 408, 584–588. https://doi.org/10.1038/35046067
Nitsche, M.A., Boggio, P.S., Fregni, F., Pascual-Leone, A., 2009. Treatment of depression with transcranial direct current stimulation (tDCS): A Review. Experimental Neurology, Brain Stimulation in Psychiatry 219, 14–19. https://doi.org/10.1016/j.expneurol.2009.03.038
Nitsche, M.A., Paulus, W., 2000. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J Physiol 527, 633–639. https://doi.org/10.1111/j.1469-7793.2000.t01-1-00633.x
Nitsche, M.A., Schauenburg, A., Lang, N., Liebetanz, D., Exner, C., Paulus, W., Tergau, F., 2003. Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. J Cogn Neurosci 15, 619–626. https://doi.org/10.1162/089892903321662994
Nitsche, M.A., Seeber, A., Frommann, K., Klein, C.C., Rochford, C., Nitsche, M.S., Fricke, K., Liebetanz, D., Lang, N., Antal, A., Paulus, W., Tergau, F., 2005. Modulating parameters of excitability during and after transcranial direct current stimulation of the human motor cortex. The Journal of Physiology 568, 291–303. https://doi.org/10.1113/jphysiol.2005.092429
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol Rev 110, 611–646. https://doi.org/10.1037/0033-295X.110.4.611
Obermayer, J., Heistek, T.S., Kerkhofs, A., Goriounova, N.A., Kroon, T., Baayen, J.C., Idema, S., Testa-Silva, G., Couey, J.J., Mansvelder, H.D., 2018. Lateral inhibition by Martinotti

interneurons is facilitated by cholinergic inputs in human and mouse neocortex. Nat Commun 9, 4101. https://doi.org/10.1038/s41467-018-06628-w
Obermayer, K., Blasdel, G.G., 1993. Geometry of orientation and ocular dominance columns in monkey striate cortex. J. Neurosci. 13, 4114–4129. https://doi.org/10.1523/JNEUROSCI.13-10-04114.1993
O'Keefe, J., Dostrovsky, J., 1971. The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. Brain Research 34, 171–175. https://doi.org/10.1016/0006-8993(71)90358-1
O'Keefe, J., Nadel, L., 1978. The hippocampus as a cognitive map. Oxford University Press, Oxford.
Okun, M., Lampl, I., 2008. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nature Neuroscience 11, 535–537. https://doi.org/10.1038/nn.2105
O'Mara, S.M., Commins, S., Anderson, M., Gigg, J., 2001. The subiculum: a review of form, physiology and function. Progress in Neurobiology 64, 129–155. https://doi.org/10.1016/S0301-0082(00)00054-X
O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proceedings of the National Academy of Sciences 110, E3660–E3669. https://doi.org/10.1073/pnas.1305373110
Öz, G., Alger, J.R., Barker, P.B., Bartha, R., Bizzi, A., Boesch, C., Bolan, P.J., Brindle, K.M., Cudalbu, C., Dinçer, A., Dydak, U., Emir, U.E., Frahm, J., González, R.G., Gruber, S., Gruetter, R., Gupta, R.K., Heerschap, A., Henning, A., Hetherington, H.P., Howe, F.A., Hüppi, P.S., Hurd, R.E., Kantarci, K., Klomp, D.W.J., Kreis, R., Kruiskamp, M.J., Leach, M.O., Lin, A.P., Luijten, P.R., Marjańska, M., Maudsley, A.A., Meyerhoff, D.J., Mountford, C.E., Nelson, S.J., Pamir, M.N., Pan, J.W., Peet, A.C., Poptani, H., Posse, S., Pouwels, P.J.W., Ratai, E.-M., Ross, B.D., Scheenen, T.W.J., Schuster, C., Smith, I.C.P., Soher, B.J., Tkáč, I., Vigneron, D.B., Kauppinen, R.A., 2014. Clinical Proton MR Spectroscopy in Central Nervous System Disorders. Radiology 270, 658–679. https://doi.org/10.1148/radiol.13130531
Öz, G., Terpstra, M., Tkáč, I., Aia, P., Lowary, J., Tuite, P.J., Gruetter, R., 2006. Proton MRS of the unilateral substantia nigra in the human brain at 4 tesla: Detection of high GABA concentrations. Magnetic Resonance in Medicine 55, 296–301. https://doi.org/10.1002/mrm.20761
Öz, G., Tkáč, I., 2011. Short-Echo , Single-Shot , Full-Intensity Proton Magnetic Resonance Spectroscopy for Neurochemical Profiling at 4 T : Validation in the Cerebellum and Brainstem. Magnetic Resonance in Medicine 65, 901–910. https://doi.org/10.1002/mrm.22708
Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H., Axmacher, N., Verschure, P.F.M.J., 2019. Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. Nature Communications 10, 2255. https://doi.org/10.1038/s41467-019-09569-0
Pastalkova, E., Itskov, V., Amarasingham, A., Buzsáki, G., 2008. Internally Generated Cell Assembly Sequences in the Rat Hippocampus. Science 321, 1322–1327. https://doi.org/10.1126/science.1159775
Pastoll, H., Solanka, L., van Rossum, M.C.W., Nolan, M.F., 2013. Feedback inhibition enables θ-nested γ oscillations and grid firing fields. Neuron 77, 141–154. https://doi.org/10.1016/j.neuron.2012.11.032
Patel, R., Spreng, R.N., Shin, L.M., Girard, T.A., 2012. Neurocircuitry models of posttraumatic stress disorder and beyond: a meta-analysis of functional neuroimaging studies. Neurosci Biobehav Rev 36, 2130–2142. https://doi.org/10.1016/j.neubiorev.2012.06.003

Pawlak, V., Wickens, J., Kirkwood, A., Kerr, J., 2010. Timing is not Everything: Neuromodulation Opens the STDP Gate. Frontiers in Synaptic Neuroscience 2.
Paz, R., Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., Fried, I., 2010. A neural substrate in the human hippocampus for linking successive events. Proceedings of the National Academy of Sciences 107, 6046–6051. https://doi.org/10.1073/pnas.0910834107
Peñagarikano, O., Abrahams, B.S., Herman, E.I., Winden, K.D., Gdalyahu, A., Dong, H., Sonnenblick, L.I., Gruver, R., Almajano, J., Bragin, A., Golshani, P., Trachtenberg, J.T., Peles, E., Geschwind, D.H., 2011. Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. Cell 147, 235–246. https://doi.org/10.1016/j.cell.2011.08.040
Peters, A., Jones, E.G. (Eds.), 1984. Cellular Components of the Cerebral Cortex. Plenum, New York, pp. 206–241.
Pfeffer, C.K., Xue, M., He, M., Huang, Z.J., Scanziani, M., 2013. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nat Neurosci 16, 1068–1076. https://doi.org/10.1038/nn.3446
Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J.I., Huang, Z.J., Kepecs, A., 2013. Cortical interneurons that specialize in disinhibitory control. Nature 503, 521–524. https://doi.org/10.1038/nature12676
Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B.V., Scanziani, M., 2009. Input normalization by global feedforward inhibition expands cortical dynamic range. Nat Neurosci 12, 1577–1585. https://doi.org/10.1038/nn.2441
Pradhan, S., Bonekamp, S., Gillen, J.S., Rowland, L.M., Wijtenburg, S.A., Edden, R.A.E., Barker, P.B., 2015. Comparison of single voxel brain MRS AT 3T and 7T using 32-channel head coils. Magnetic Resonance Imaging 33, 1013–1018. https://doi.org/10.1016/j.mri.2015.06.003
Preston, A.R., Eichenbaum, H., 2013. Interplay of hippocampus and prefrontal cortex in memory. Current biology 23, R764–R773. https://doi.org/10.1016/j.cub.2013.05.041
Preuschoff, K., 't Hart, B., Einhauser, W., 2011. Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Frontiers in Neuroscience 5.
Prinsen, H., Graaf, R.A. de, Mason, G.F., Pelletier, D., Juchem, C., 2017. Reproducibility measurement of glutathione, GABA, and glutamate: Towards in vivo neurochemical profiling of multiple sclerosis with MR spectroscopy at 7T. Journal of Magnetic Resonance Imaging 45, 187–198. https://doi.org/10.1002/jmri.25356
Provencher, S.W., 2001. Automatic quantitation of localized in vivo 1H spectra with LCModel. NMR Biomed 14, 260–264. https://doi.org/10.1002/nbm.698
Provencher, S.W., 1993. Estimation of Metabolite Concentrations from Localized in Vivo Proton NMR Spectra. Magnetic Resonance in Medicine 30, 672–679.
Purpura, D.P., McMurtry, J.G., 1965. Intracellular activities and evoked potential changes during polarization of motor cortex. Journal of Neurophysiology 28, 166–185. https://doi.org/10.1152/jn.1965.28.1.166
Puts, N.A.J., Edden, R.A.E., 2012. In vivo magnetic resonance spectroscopy of GABA: A methodological review. Prog Nucl Magn Reson Spectrosc 60, 29–41. https://doi.org/10.1016/j.pnmrs.2011.06.001
Puts, N.A.J., Edden, R.A.E., Evans, C.J., McGlone, F., McGonigle, D.J., 2011. Regionally Specific Human GABA Concentration Correlates with Tactile Discrimination Thresholds. J. Neurosci. 31, 16556–16560. https://doi.org/10.1523/JNEUROSCI.4489-11.2011
Rae, C., 2014. A Guide to the Metabolic Pathways and Function of Metabolites Observed in Human Brain 1H Magnetic Resonance Spectra. Neurochemical Research 39, 1–36.

Ragozzino, M.E., Kim, J., Hassert, D., Minniti, N., Kiang, C., 2003. The contribution of the rat prelimbic-infralimbic areas to different forms of task switching. Behav Neurosci 117, 1054–1065. https://doi.org/10.1037/0735-7044.117.5.1054

Raiteri, M., Del Carmine, R., Cervoni, A.M., Levi, G., 1979. Differential binding of antidepressants to noradrenaline and serotonin transport sites in central nerve endings. European Journal of Pharmacology 57, 407–416. https://doi.org/10.1016/0014-2999(79)90503-X

Rajkowski, J., Kubiak, P., Aston-Jones, G., 1993. Correlations between locus coeruleus (LC) neural activity, pupil diameter and behavior in monkey support a role of LC in attention. Society for Neuroscience Abstracts.

Reis, J., Schambra, H.M., Cohen, L.G., Buch, E.R., Fritsch, B., Zarahn, E., Celnik, P.A., Krakauer, J.W., 2009. Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. Proc Natl Acad Sci U S A 106, 1590–1595. https://doi.org/10.1073/pnas.0805413106

Ribak, C.E., Seress, L., Amaral, D.G., 1985. The development, ultrastructure and synaptic connections of the mossy cells of the dentate gyrus. J Neurocytol 14, 835–857. https://doi.org/10.1007/BF01170832

Rich, E.L., Shapiro, M.L., 2007. Prelimbic/infralimbic inactivation impairs memory for multiple task switches, but not flexible selection of familiar tasks. J Neurosci 27, 4747–4755. https://doi.org/10.1523/jneurosci.0369-07.2007

Riches, I.P., Wilson, F.A., Brown, M.W., 1991. The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. J Neurosci 11, 1763–1779.

Rideaux, R., 2021. No balance between glutamate+glutamine and GABA+ in visual or motor cortices of the human brain: A magnetic resonance spectroscopy study. NeuroImage 237, 118191. https://doi.org/10.1016/j.neuroimage.2021.118191

Robertson, C.E., Ratai, E.-M., Kanwisher, N., 2016. Reduced GABAergic Action in the Autistic Brain. Current Biology 26, 80–85. https://doi.org/10.1016/j.cub.2015.11.019

Rolls, E.T., 2008. Memory, Attention, and Decision-Making: A Unifying Computational Neuroscience Approach. Oxford University Press, Oxford. https://doi.org/10.1093/acprof:oso/9780199232703.001.0001

Rosenkranz, K., Kacar, A., Rothwell, J.C., 2007. Differential Modulation of Motor Cortical Plasticity and Excitability in Early and Late Phases of Human Motor Learning. J. Neurosci. 27, 12058–12066. https://doi.org/10.1523/JNEUROSCI.2663-07.2007

Rothman, D.L., Behar, K.L., Hyder, F., Shulman, R.G., 2003. In vivo NMR Studies of the Glutamate Neurotransmitter Flux and Neuroenergetics: Implications for Brain Function. Annual Review of Physiology 65, 401–427. https://doi.org/10.1146/annurev.physiol.65.092101.142131

Rothman, D.L., De Feyter, H.M., de Graaf, R.A., Mason, G.F., Behar, K.L., 2011. 13C MRS studies of neuroenergetics and neurotransmitter cycling in humans. NMR in Biomedicine 24, 943–957. https://doi.org/10.1002/nbm.1772

Royer, S., Zemelman, B.V., Losonczy, A., Kim, J., Chance, F., Magee, J.C., Buzsáki, G., 2012. Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. Nat Neurosci 15, 769–775. https://doi.org/10.1038/nn.3077

Rudy, B., Fishell, G., Lee, S., Hjerling-Leffler, J., 2011. Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Developmental Neurobiology 71, 45–61. https://doi.org/10.1002/dneu.20853

Sampaio-Baptista, C., Filippini, N., Stagg, C.J., Near, J., Scholz, J., Johansen-Berg, H., 2015. Changes in functional connectivity and GABA levels with long-term motor learning. Neuroimage 106, 15–20. https://doi.org/10.1016/j.neuroimage.2014.11.032

Samuels, E.R., Szabadi, E., 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part II: physiological and pharmacological manipulations and pathological alterations of locus coeruleus activity in humans. Curr Neuropharmacol 6, 254–285. https://doi.org/10.2174/157015908785777193
Sanaei Nezhad, F., Anton, A., Parkes, L.M., Deakin, B., Williams, S.R., 2017. Quantification of glutathione in the human brain by MR spectroscopy at 3 Tesla: Comparison of PRESS and MEGA-PRESS. Magnetic Resonance in Medicine 78, 1257–1266. https://doi.org/10.1002/mrm.26532
Sanchez-Vives, M.V., McCormick, D.A., 2000. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. Nat Neurosci 3, 1027–1034. https://doi.org/10.1038/79848
Sanger, D.J., Joly, D., 1985. Anxiolytic drugs and the acquisition of conditioned fear in mice. Psychopharmacology 85, 284–288. https://doi.org/10.1007/BF00428188
Sapountzis, P., Schluppeck, D., Bowtell, R., Peirce, J.W., 2010. A comparison of fMRI adaptation and multivariate pattern classification analysis in visual cortex. Neuroimage 49, 1632–1640. https://doi.org/10.1016/j.neuroimage.2009.09.066
Sara, S.J., 2009. The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 10, 211–223. https://doi.org/10.1038/nrn2573
Sara, S.J., Bouret, S., 2012. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. Neuron 76, 130–141. https://doi.org/10.1016/j.neuron.2012.09.011
Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B.L., Witter, M.P., Moser, M.-B., Moser, E.I., 2006. Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex. Science 312, 758–762. https://doi.org/10.1126/science.1125572
Sauer, J.-M., Ring, B.J., Witcher, J.W., 2005. Clinical Pharmacokinetics of Atomoxetine. Clin Pharmacokinet 44, 571–590. https://doi.org/10.2165/00003088-200544060-00002
Sawamura, H., Orban, G.A., Vogels, R., 2006. Selectivity of Neuronal Adaptation Does Not Match Response Selectivity: A Single-Cell Study of the fMRI Adaptation Paradigm. Neuron 49, 307–318. https://doi.org/10.1016/j.neuron.2005.11.028
Schapiro, A.C., Kustner, L.V., Turk-Browne, N.B., 2012. Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. Current Biology 22, 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056
Schapiro, A.C., Rogers, T.T., Cordova, N.I., Turk-Browne, N.B., Botvinick, M.M., 2013. Neural representations of events arise from temporal community structure. Nat Neurosci 16, 486–492. https://doi.org/10.1038/nn.3331
Schlichting, M.L., Preston, A.R., 2014. Memory reactivation during rest supports upcoming learning of related content. Proceedings of the National Academy of Sciences 111, 15845–15850. https://doi.org/10.1073/pnas.1404396111
Schmitz, T.W., Correia, M.M., Ferreira, C.S., Prescot, A.P., Anderson, M.C., 2017. Hippocampal GABA enables inhibitory control over unwanted thoughts. Nat Commun 8, 1311. https://doi.org/10.1038/s41467-017-00956-z
Scholl, J., Kolling, N., Nelissen, N., Stagg, C.J., Harmer, C.J., Rushworth, M.F., 2017. Excitation and inhibition in anterior cingulate predict use of past experiences. eLife 6. https://doi.org/10.7554/eLife.20365
Scoville, W.B., Milner, B., 1957. Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry 20, 11–21.
Seamans, J.K., Yang, C.R., 2004. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog Neurobiol 74, 1–58. https://doi.org/10.1016/j.pneurobio.2004.05.006
Shapiro, M.L., Olton, D.S., 1994. Hippocampal function and interference, in: Memory Systems 1994. The MIT Press, Cambridge, MA, US, pp. 87–117.

Shen, J., Petersen, K.F., Behar, K.L., Brown, P., Nixon, T.W., Mason, G.F., Petroff, O.A.C., Shulman, G.I., Shulman, R.G., Rothman, D.L., 1999. Determination of the rate of the glutamate/glutamine cycle in the human brain by in vivo 13C NMR. PNAS 96, 8235–8240. https://doi.org/10.1073/pnas.96.14.8235

Sherry, D.F., Schacter, D.L., 1987. The evolution of multiple memory systems. Psychological Review 94, 439–454. https://doi.org/10.1037/0033-295X.94.4.439

Shibata, K., Sasaki, Y., Bang, J.W., Walsh, E.G., Machizawa, M.G., Tamaki, M., Chang, L., Watanabe, T., 2017. Overlearning hyperstabilizes a skill by rapidly making neurochemical processing inhibitory-dominant. Nature Neuroscience 20, 470–475. https://doi.org/10.1038/nn.4490

Siapas, A.G., Lubenov, E.V., Wilson, M.A., 2005. Prefrontal Phase Locking to Hippocampal Theta Oscillations. Neuron 46, 141–151. https://doi.org/10.1016/j.neuron.2005.02.028

Sibson, N.R., Dhankhar, A., Mason, G.F., Rothman, D.L., Behar, K.L., Shulman, R.G., 1998. Stoichiometric coupling of brain glucose metabolism and glutamatergic neuronal activity. PNAS 95, 316–321. https://doi.org/10.1073/pnas.95.1.316

Simpson, D., Perry, C.M., 2003. Atomoxetine. Pediatr-Drugs 5, 407–415. https://doi.org/10.2165/00128072-200305060-00005

Singer, A.C., Carr, M.F., Karlsson, M.P., Frank, L.M., 2013. Hippocampal SWR activity predicts correct decisions during the initial learning of an alternation task. Neuron 77, 1163–1173. https://doi.org/10.1016/j.neuron.2013.01.027

Sohal, V.S., Zhang, F., Yizhar, O., Deisseroth, K., 2009. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. Nature 459, 698–702. https://doi.org/10.1038/nature07991

Song, S., Abbott, L.F., 2001. Cortical development and remapping through spike timing-dependent plasticity. Neuron 32, 339–350. https://doi.org/10.1016/s0896-6273(01)00451-2

Sparta, D.R., Hovelsø, N., Mason, A.O., Kantak, P.A., Ung, R.L., Decot, H.K., Stuber, G.D., 2014. Activation of Prefrontal Cortical Parvalbumin Interneurons Facilitates Extinction of Reward-Seeking Behavior. J Neurosci 34, 3699–3705. https://doi.org/10.1523/JNEUROSCI.0235-13.2014

Squire, Larry R., 1992. Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory. Journal of Cognitive Neuroscience 4, 232–243. https://doi.org/10.1162/jocn.1992.4.3.232

Squire, L.R., 1992. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review 99(2), 195–231.

Stachenfeld, K.L., Botvinick, M.M., Gershman, S.J., 2017. The hippocampus as a predictive map. Nat Neurosci 20, 1643–1653. https://doi.org/10.1038/nn.4650

Stagg, C., Rothman, D.L., 2013. Magnetic Resonance Spectroscopy: Tools for Neuroscience Research and Emerging Clinical Applications. Academic Press.

Stagg, C.J., Bachtiar, V., Johansen-Berg, H., 2011. The role of GABA in human motor learning. Current biology 21, 480–484. https://doi.org/10.1016/j.cub.2011.01.069

Stagg, C.J., Best, J.G., Stephenson, M.C., O'Shea, J., Wylezinska, M., Kincses, Z.T., Morris, P.G., Matthews, P.M., Johansen-Berg, H., 2009. Polarity-Sensitive Modulation of Cortical Neurotransmitters by Transcranial Stimulation. J. Neurosci. 29, 5202–5206. https://doi.org/10.1523/JNEUROSCI.4432-08.2009

Stagg, C.J., Nitsche, M.A., 2011. Physiological Basis of Transcranial Direct Current Stimulation. Neuroscientist 17, 37–53. https://doi.org/10.1177/1073858410386614

Stanley, J.A., Raz, N., 2018. Functional Magnetic Resonance Spectroscopy: The "New" MRS for Cognitive Neuroscience and Psychiatry Research. Front. Psychiatry 9. https://doi.org/10.3389/fpsyt.2018.00076

Steemers, B., Vicente-Grabovetsky, A., Barry, C., Smulders, P., Schröder, T.N., Burgess, N., Doeller, C.F., 2016. Hippocampal Attractor Dynamics Predict Memory-Based Decision Making. Curr Biol 26, 1750–1757. https://doi.org/10.1016/j.cub.2016.04.063

Stemmler, M., Mathis, A., Herz, A.V.M., 2015. Connecting multiple spatial scales to decode the population activity of grid cells. Science Advances 1, e1500816. https://doi.org/10.1126/science.1500816

Strange, B.A., Hurlemann, R., Dolan, R.J., 2003. An emotion-induced retrograde amnesia in humans is amygdala- and β-adrenergic-dependent. Proceedings of the National Academy of Sciences 100, 13626–13631. https://doi.org/10.1073/pnas.1635116100

Stringer, C., Pachitariu, M., Steinmetz, N.A., Okun, M., Bartho, P., Harris, K.D., Sahani, M., Lesica, N.A., 2016. Inhibitory control of correlated intrinsic variability in cortical networks. eLife 5, e19695. https://doi.org/10.7554/eLife.19695

Sugisaki, E., Fukushima, Y., Tsukada, M., Aihara, T., 2011. Cholinergic modulation on spike timing-dependent plasticity in hippocampal CA1 network. Neuroscience 192, 91–101. https://doi.org/10.1016/j.neuroscience.2011.06.064

Summerfield, C., Trittschuh, E.H., Monti, J.M., Mesulam, M.-M., Egner, T., 2008. Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11, 1004–1006. https://doi.org/10.1038/nn.2163

Sutherland, R.J., Rudy, J.W., 1989. Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. Psychobiology 17, 129–144. https://doi.org/10.3758/BF03337828

Swanson, C.J., Perry, K.W., Koch-Krueger, S., Katner, J., Svensson, K.A., Bymaster, F.P., 2006. Effect of the attention deficit/hyperactivity disorder drug atomoxetine on extracellular concentrations of norepinephrine and dopamine in several brain regions of the rat. Neuropharmacology 50, 755–760. https://doi.org/10.1016/j.neuropharm.2005.11.022

Swanson, L.W., Kohler, C., 1986. Anatomical evidence for direct projections from the entorhinal area to the entire cortical mantle in the rat. J. Neurosci. 6, 3010–3023. https://doi.org/10.1523/JNEUROSCI.06-10-03010.1986

Takata, N., Sugiura, Y., Yoshida, K., Koizumi, M., Hiroshi, N., Honda, K., Yano, R., Komaki, Y., Matsui, K., Suematsu, M., Mimura, M., Okano, H., Tanaka, K.F., 2018. Optogenetic astrocyte activation evokes BOLD fMRI response with oxygen consumption without neuronal activity modulation. Glia 66, 2013–2023. https://doi.org/10.1002/glia.23454

Tanaka, K.Z., Pevzner, A., Hamidi, A.B., Nakazawa, Y., Graham, J., Wiltgen, B.J., 2014. Cortical Representations Are Reinstated by the Hippocampus during Memory Retrieval. Neuron 84, 347–354. https://doi.org/10.1016/j.neuron.2014.09.037

Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsani, D., Fu, Y., Lu, J., Lin, Y., Miyoshi, G., Shima, Y., Fishell, G., Nelson, S.B., Huang, Z.J., 2011. A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex. Neuron 71, 995–1013. https://doi.org/10.1016/j.neuron.2011.07.026

Tavares, R.M., Mendelsohn, A., Grossman, Y., Williams, C.H., Shapiro, M., Trope, Y., Schiller, D., 2015. A Map for Social Navigation in the Human Brain. Neuron 87, 231–243. https://doi.org/10.1016/j.neuron.2015.06.011

Taylor, R., Neufeld, R.W.J., Schaefer, B., Densmore, M., Rajakumar, N., Osuch, E.A., Williamson, P.C., Théberge, J., 2015. Functional magnetic resonance spectroscopy of glutamate in schizophrenia and major depressive disorder: anterior cingulate activity during a color-word Stroop task. npj Schizophrenia 1, 1–8. https://doi.org/10.1038/npjschz.2015.28

Teeuwisse, W.M., Brink, W.M., Webb, A.G., 2012. Quantitative assessment of the effects of high-permittivity pads in 7 Tesla MRI of the brain. Magn Reson Med 67, 1285–1293. https://doi.org/10.1002/mrm.23108

Terpstra, M., Cheong, I., Lyu, T., Deelchand, D.K., Emir, U.E., Bednařík, P., Eberly, L.E., Öz, G., 2016. Test-retest reproducibility of neurochemical profiles with short-echo, single-voxel MR spectroscopy at 3T and 7T. Magnetic Resonance in Medicine 76, 1083–1091. https://doi.org/10.1002/mrm.26022

Terpstra, M., Marjanska, M., Henry, P.-G., Tkáč, I., Gruetter, R., 2006. Detection of an antioxidant profile in the human brain in vivo via double editing with MEGA-PRESS. Magnetic Resonance in Medicine 56, 1192–1199. https://doi.org/10.1002/mrm.21086

Teyler, T.J., DiScenna, P., 1985. The role of hippocampus in memory: A hypothesis. Neuroscience & Biobehavioral Reviews 9, 377–389. https://doi.org/10.1016/0149-7634(85)90016-8

Teyler, T.J., Rudy, J.W., 2007. The hippocampal indexing theory and episodic memory: updating the index. Hippocampus 17, 1158–1169. https://doi.org/10.1002/hipo.20350

Thompson, R.F., 2005. In Search of Memory Traces. Annual Review of Psychology 56, 1–23. https://doi.org/10.1146/annurev.psych.56.091103.070239

Tkáč, I., 2008. Refinement of simulated basis set for LCModel analysis. Presented at the Proc. 16th Annu. Meet. ISMRM Toronto Canada.

Tkáč, I., Andersen, P., Adriany, G., Merkle, H., Uğurbil, K., Gruetter, R., 2001. In vivo 1H NMR spectroscopy of the human brain at 7 T. Magnetic Resonance in Medicine 46, 451–456. https://doi.org/10.1002/mrm.1213

Tkáč, I., Starčuk, Z., Choi, I.-Y., Gruetter, R., 1999. In vivo 1H NMR spectroscopy of rat brain at 1 ms echo time. Magnetic Resonance in Medicine 41, 649–656. https://doi.org/10.1002/(SICI)1522-2594(199904)41:4<649::AID-MRM2>3.0.CO;2-G

Todorovic, A., De Lange, F.P., 2012. Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. J. Neurosci. 32, 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Tolman, E.C., 1948. Cognitive maps in rats and men. Psychological Review 55, 189–208. https://doi.org/10.1037/h0061626

Trabesinger, A.H., Boesiger, P., 2001. Improved selectivity of double quantum coherence filtering for the detection of glutathione in the human brain in vivo. Magnetic Resonance in Medicine 45, 708–710. https://doi.org/10.1002/mrm.1094

Tremblay, R., Lee, S., Rudy, B., 2016. GABAergic interneurons in the neocortex: From cellular properties to circuits. Neuron 91, 260–292. https://doi.org/10.1016/j.neuron.2016.06.033

Trepel, C., Racine, R.J., 2000. GABAergic modulation of neocortical long-term potentiation in the freely moving rat. Synapse 35, 120–128. https://doi.org/10.1002/(SICI)1098-2396(200002)35:2<120::AID-SYN4>3.0.CO;2-6

Treves, A., Rolls, E.T., 1994. Computational analysis of the role of the hippocampus in memory. Hippocampus 4, 374–391. https://doi.org/10.1002/hipo.450040319

Tronel, S., Belnoue, L., Grosjean, N., Revest, J.-M., Piazza, P.-V., Koehl, M., Abrous, D.N., 2012. Adult-born neurons are necessary for extended contextual discrimination. Hippocampus 22, 292–298. https://doi.org/10.1002/hipo.20895

Tse, D., Langston, R.F., Kakeyama, M., Bethus, I., Spooner, P.A., Wood, E.R., Witter, M.P., Morris, R.G.M., 2007. Schemas and Memory Consolidation. Science 316, 76–82. https://doi.org/10.1126/science.1135935

Tulving, E., 1974. Cue-Dependent Forgetting: When we forget something we once knew, it does not necessarily mean that the memory trace has been lost; it may only be inaccessible. American Scientist 62, 74–82.

Turrigiano, G.G., 2008. The Self-Tuning Neuron: Synaptic Scaling of Excitatory Synapses. Cell 135, 422–435. https://doi.org/10.1016/j.cell.2008.10.008

Tzavara, E.T., Bymaster, F.P., Overshiner, C.D., Davis, R.J., Perry, K.W., Wolff, M., McKinzie, D.L., Witkin, J.M., Nomikos, G.G., 2006. Procholinergic and memory enhancing

properties of the selective norepinephrine uptake inhibitor atomoxetine. Mol Psychiatry 11, 187–195. https://doi.org/10.1038/sj.mp.4001763
Urakubo, H., Honda, M., Froemke, R.C., Kuroda, S., 2008. Requirement of an Allosteric Kinetics of NMDA Receptors for Spike Timing-Dependent Plasticity. J. Neurosci. 28, 3310–3323. https://doi.org/10.1523/JNEUROSCI.0303-08.2008
Urban-Ciecko, J., Barth, A.L., 2016. Somatostatin-expressing neurons in cortical networks. Nat Rev Neurosci 17, 401–409. https://doi.org/10.1038/nrn.2016.53
Urban-Ciecko, J., Fanselow, E.E., Barth, A.L., 2015. Neocortical somatostatin neurons reversibly silence excitatory transmission via GABAb receptors. Curr Biol 25, 722–731. https://doi.org/10.1016/j.cub.2015.01.035
Vaiva, G., Thomas, P., Ducrocq, F., Fontaine, M., Boss, V., Devos, P., Rascle, C., Cottencin, O., Brunet, A., Laffargue, P., Goudemand, M., 2004. Low posttrauma GABA plasma levels as a predictive factor in the development of acute posttraumatic stress disorder. Biol Psychiatry 55, 250–254. https://doi.org/10.1016/j.biopsych.2003.08.009
Vallentin, D., Kosche, G., Lipkind, D., Long, M.A., 2016. Inhibition protects acquired song segments during vocal learning in zebra finches. Science 351, 267–271. https://doi.org/10.1126/science.aad3023
van Praag, H., Schinder, A.F., Christie, B.R., Toni, N., Palmer, T.D., Gage, F.H., 2002. Functional neurogenesis in the adult hippocampus. Nature 415, 1030–1034. https://doi.org/10.1038/4151030a
Vaz, A.P., Inati, S.K., Brunel, N., Zaghloul, K.A., 2019. Coupled ripple oscillations between the medial temporal lobe and neocortex retrieve human memory. Science 363, 975–978. https://doi.org/10.1126/science.aau8956
Verhagen, L., Gallea, C., Folloni, D., Constans, C., Jensen, D.E., Ahnine, H., Roumazeilles, L., Santin, M., Ahmed, B., Lehericy, S., Klein-Flügge, M.C., Krug, K., Mars, R.B., Rushworth, M.F., Pouget, P., Aubry, J.-F., Sallet, J., 2019. Offline impact of transcranial focused ultrasound on cortical activation in primates. eLife 8, e40541. https://doi.org/10.7554/eLife.40541
Vogels, T.P., Abbott, L.F., 2009. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. Nat Neurosci 12, 483–491. https://doi.org/10.1038/nn.2276
Vogels, T.P., Abbott, L.F., 2007. Gating deficits in model networks: a path to schizophrenia? Pharmacopsychiatry 40 Suppl 1, S73-77. https://doi.org/10.1055/s-2007-992130
Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., Gerstner, W., 2011. Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. Science 334, 1569–1573. https://doi.org/10.1126/science.1211095.
Wagner, J.J., Alger, B.E., 1996. Homosynaptic LTD and depotentiation: Do they differ in name only? Hippocampus 6, 24–29. https://doi.org/10.1002/(SICI)1098-1063(1996)6:1<24::AID-HIPO5>3.0.CO;2-7
Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., Diedrichsen, J., 2016. Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage 137, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012
Wang, S.-H., Morris, R.G.M., 2010. Hippocampal-Neocortical Interactions in Memory Formation, Consolidation, and Reconsolidation. Annual Review of Psychology 61, 49–79. https://doi.org/10.1146/annurev.psych.093008.100523
Warren, W.H., Rothman, D.B., Schnapp, B.H., Ericson, J.D., 2017. Wormholes in virtual space: From cognitive maps to cognitive graphs. Cognition 166, 152–163. https://doi.org/10.1016/j.cognition.2017.05.020
Wehr, M., Zador, A.M., 2003. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426(6965), 442–446.

Weinrich, C.A., Brittain, J.-S., Nowak, M., Salimi-Khorshidi, R., Brown, P., Stagg, C.J., 2017. Modulation of Long-Range Connectivity Patterns via Frequency-Specific Stimulation of Human Cortex. Current Biology 27, 3061-3068.e3. https://doi.org/10.1016/j.cub.2017.08.075
Wheeler, M.E., Petersen, S.E., Buckner, R.L., 2000. Memory's echo: Vivid remembering reactivates sensory-specific cortex. PNAS 97, 11125–11129. https://doi.org/10.1073/pnas.97.20.11125
White, E.L., 1989. Cortical Circuits. Synaptic Organization of the Cerebral Cortex. Birkhauser, Boston.
Whittington, J.C.R., Muller, T.H., Mark, S., Chen, G., Barry, C., Burgess, N., Behrens, T.E.J., 2020. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. Cell 183, 1249-1263.e23. https://doi.org/10.1016/j.cell.2020.10.024
Wiestler, T., McGonigle, D.J., Diedrichsen, J., 2011. Integration of sensory and motor representations of single fingers in the human cerebellum. Journal of Neurophysiology 105, 3042–3053. https://doi.org/10.1152/jn.00106.2011
Wiggs, C.L., Martin, A., 1998. Properties and mechanisms of perceptual priming. Current Opinion in Neurobiology 8, 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X
Wikenheiser, A.M., Schoenbaum, G., 2016. Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. Nat Rev Neurosci 17, 513–523. https://doi.org/10.1038/nrn.2016.56
Wilent, W.B., Contreras, D., 2005. Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. Nat Neurosci 8, 1364–1370. https://doi.org/10.1038/nn1545
Wills, T.J., Cacucci, F., Burgess, N., O'Keefe, J., 2010. Development of the Hippocampal Cognitive Map in Preweanling Rats. Science 328, 1573–1576. https://doi.org/10.1126/science.1188224
Wilson, M.A., McNaughton, B.L., 1994. Reactivation of Hippocampal Ensemble Memories During Sleep. Science 265, 676–679. https://doi.org/10.1126/science.8036517
Wimmer, G.E., Shohamy, D., 2012. Preference by Association: How Memory Mechanisms in the Hippocampus Bias Decisions. Science 338, 270–273.
Wolff, S.B.E., Gründemann, J., Tovote, P., Krabbe, S., Jacobson, G.A., Müller, C., Herry, C., Ehrlich, I., Friedrich, R.W., Letzkus, J.J., Lüthi, A., 2014. Amygdala interneuron subtypes control fear learning through disinhibition. Nature 509, 453–458. https://doi.org/10.1038/nature13258
Wong, D.T., Threlkeld, P.G., Best, K.L., Bymaster, F.P., 1982. A new inhibitor of norepinephrine uptake devoid of affinity for receptors in rat brain. J Pharmacol Exp Ther 222, 61–65.
Woody, C.D., Gruen, E., 1987. Acetylcholine reduces net outward currents measured in vivo with single electrode voltage clamp techniques in neurons of the motor cortex of cats. Brain Research 424, 193–198. https://doi.org/10.1016/0006-8993(87)91210-8
Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited--again. Neuroimage 2, 173–181. https://doi.org/10.1006/nimg.1995.1023
Xiang, Z., Huguenard, J.R., Prince, D.A., 1998. Cholinergic Switching Within Neocortical Inhibitory Networks. Science 281, 985–988. https://doi.org/10.1126/science.281.5379.985
Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J.A., Poldrack, R.A., 2010. Greater Neural Pattern Similarity Across Repetitions Is Associated with Better Memory. Science 330, 97–101. https://doi.org/10.1126/science.1193125
Xue, M., Atallah, B.V., Scanziani, M., 2014. Equalizing excitation–inhibition ratios across visual cortical neurons. Nature 511, 596–600. https://doi.org/10.1038/nature13321

Yang, S.-N., Tang, Y.-G., Zucker, R.S., 1999. Selective Induction of LTP and LTD by Postsynaptic [Ca $^{2+}$] $_{i}$ Elevation. Journal of Neurophysiology 81, 781–787. https://doi.org/10.1152/jn.1999.81.2.781

Yassa, M.A., Stark, C.E.L., 2011. Pattern separation in the hippocampus. Trends in Neurosciences, Special Issue: Hippocampus and Memory 34, 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yavorska, I., Wehr, M., 2016. Somatostatin-Expressing Inhibitory Interneurons in Cortical Circuits. Frontiers in Neural Circuits 10.

Yehuda, R., McFarlane, A.C., Shalev, A.Y., 1998. Predicting the development of posttraumatic stress disorder from the acute response to a traumatic event. Biol Psychiatry 44, 1305–1313. https://doi.org/10.1016/s0006-3223(98)00276-5

Yizhar, O., Fenno, L.E., Prigge, M., Schneider, F., Davidson, T.J., O'Shea, D.J., Sohal, V.S., Goshen, I., Finkelstein, J., Paz, J.T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J.R., Hegemann, P., Deisseroth, K., 2011. Neocortical excitation/inhibition balance in information processing and social dysfunction. Nature 477, 171–178. https://doi.org/10.1038/nature10360

Yu, A.J., Dayan, P., 2005. Uncertainty, Neuromodulation, and Attention. Neuron 46, 681–692. https://doi.org/10.1016/j.neuron.2005.04.026

Zamoscik, V., Mier, D., Schmidt, S.N.L., Kirsch, P., 2016. Early Memories of Individuals on the Autism Spectrum Assessed Using Online Self-Reports. Frontiers in Psychiatry 7.

Zenke, F., Agnes, E.J., Gerstner, W., 2015. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nat Commun 6, 6922. https://doi.org/10.1038/ncomms7922

Zhang, S., Xu, M., Kamigaki, T., Do, J.P.H., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L., Dan, Y., 2014. Long-range and local circuits for top-down modulation of visual cortex processing.